DOCUMENT RESUME

ED 101 757 IR 001 717

AUTHOR	Bourne, Charles P.; And Others
TITLE	Analysis of ERIC On-Line File Searching Procedures and Guidelines for Searching. Final Report.
INSTITUTION	California Univ., Berkeley. Inst. of Library Research.; Lockheed Research Lab., Palo Alto, Calif.
SPONS AGENCY	National Inst. of Education (DHEW), Washington, D.C.
REPORT NO	ILR-74-005
PUB DATE	Nov 74
CONTRACT	OEC-0-71-2559
NOTE	156p.
EDRS PRICE	MF-$0.76 HC-$8.24 PLUS POSTAGE
DESCRIPTORS	*Data Bases; Electronic Equipment; *Information Retrieval; Information Systems; Input Output Devices; Library Research; *On Line Systems; Relevance (Information Retrieval); *Search Strategies; *Thesauri; Use Studies
IDENTIFIERS	*DIALOG; Educational Resources Information Center; ERIC

ABSTRACT

In order to identify and explain the reasons for differences in average on-line search rates among terminal installations, the operations and performance of several facilities using the Lockheed DIALOG system for on-line searching of the ERIC data base were studied. Detailed examinations were made of such aspects as the DIALOG system response time as a function of the time of day or day of the week; the search commands and logic used by each of the terminal installations for their operations; the mix of complex, medium, or simple questions processed at each terminal location; and the extent and impact of the variant forms of descriptors in the file (e.g., singular and plural forms of the same term). Timing studies were performed to suggest some terminal procedures that could increase average on-line search speeds. Guidelines for searchers to consider for pre-search and terminal activities are presented at the end of the study. (Author/PF)

ED101757

BEST COPY AVAILABLE

ILR-74-005

Final Report

Analysis of ERIC On-line File Searching Procedures and Guidelines for Searching

Charles P. Bourne
Jo Robinson
Judy Todd

Institute of Library Research
University of California

Berkeley, California 94720

November 1974

Performed for Lockheed Palo Alto Research Laboratory, Palo Alto, California, under their contract No. OEC-0-71-2559 to the U. S. National Institute of Education

IR 001717

ED101757

ILR-74-005

Final Report

Analysis of ERIC On-line File Searching Procedures and Guidelines for Searching

Charles P. Bourne
Jo Robinson
Judy Todd

Institute of Library Research
University of California

Berkeley, California 94720

November 1974

Performed for Lockheed Palo Alto Research Laboratory, Palo Alto, California, under their contract No. OEC-0-71-2559 to the U. S. National Institute of Education

CONTENTS

Page

ABSTRACT 1

FIGURES 2-3

TABLES 4

ACKNOWLEDGMENTS 5

I. INTRODUCTION 6

A. The Problem 6

B. Objectives 9

II. SUMMARY 10-12

III. REVIEW OF EXISTING LITERATURE 13

A. Important Monographs 13

B. Bibliographies and Reviews 14

C. Prior Evaluation Studies 15

D. Other Studies 17

E. Other Work in Progress 18

F. References 19-31

IV. BACKGROUND INFORMATION 32

A. The ERIC Data Base 32

B. Searching the ERIC Data Base By Computer 33

C. Lockheed DIALOG System 34-38

D. References 39

V. ANALYSIS OF TERMINAL INSTALLATIONS 40

A. Installations Studied 40

B. Local Characteristics 41-43

VI. ANALYSIS OF ERIC/DIALOG USE 44

A. Introduction 44

B. Study of ERIC/DIALOG System Response Time 45

CONTENTS (cont'd)

Page

1. Introduction 45
2. Methodology 45-47
3. Possible Sources of Methodological Error 47-48
4. Results 48-52
5. Conclusions 52
    a) Influence of Search Operators 52
    b) Influence of Hour of the Day 52-59
    c) Influence of Day of the Week 59

C. Detailed Examination of ERIC Searches 60
1. The Source Data 60
2. Definition of Search and Question 60-63
3. Findings 63-75

D. Classification of ERIC Questions 76
1. Classification Rules 76-78
2. The Data 79
    a) Search Time as a Function of Question Complexity 79
    b) Question Complexity as a Function of the Installation 79
    c) Question Complexity as a Function of the Terminal Equipment 79
    d) Question Mix as a Function of Personal Work Habits of the Analysts 79-82

E. Variant Forms of ERIC Descriptors and Identifiers 83
1. Background 83
2. Nature and Extent of Variant Forms 83-88
3. Impact of Variant Forms on Search Output 89-90

CONTENTS (cont'd)

Page

F. Influence of Other Factors on Search Speeds 91
  1. Terminal Equipment 91
  2. Continuing Education, and Association with Other Searchers 91-93
  3. Subject Expertise 93
  4. Extent of Pre-planning before Searching 93
  5. Fee Versus Free Service, and Cost-Conscious Attitudes 93-94
  6. User Versus Intermediary Searching, and Extent of User Involvement 94
  7. Availability and Use of Analyst Support Tools 94-95
G. References 96

VII. GUIDELINES FOR SEARCHING THE ERIC FILES USING DIALOG 97

A. Introduction 97-98
B. Pre-Search Activity 99
  1. Decision: Whether to Go On-Line 99-103
  2. Decision: Whether to Use Printed Aids Such As the ERIC Thesaurus Before Going to the Terminal 104
    a) The ERIC Thesaurus 104-106
    b) Term Frequency Lists 106-108
    c) Other Printed Aids 109
  3. Decision: Whether to Use SEARCH SAVE 109-111
  4. Advance Determination of Possible Ways to Limit Output 111
C. Terminal Activity 112
  1. Equipment Considerations 112-113
  2. Recommended Keyboard Procedures 113-114
    a) Initialization 114-115
    b) Relative Merits of EXPAND-SELECT Combination vs. SELECT Alone 115-120

CONTENTS (cont'd)

Page

c) Relative Merits of SELECT and Straight Typing vs. SELECT E- or R- Numbers 120-122

d) Timing of Five Alternative Procedures for SELECTing Descriptors 122-124

3. Strategy 124

a) Number of Terms Needed to Adequately Express Each Facet of a Multi-Facet Search 124-133

b) Methods of Limiting Quantity of Output, and Their Effect on Relevance 133

1) Limit by LIMIT/MAJ Command 133-135

2) Limit by Accession Number 135

3) Limit by Printing Only the First N Citations 135

4) Results of All Attempts to Limit Output 135-138

D. References 139-140

APPENDIX A -- LOG OF TIMING DATA FOR 7-DAY TIMING TEST

APPENDIX B -- INCREMENTAL SEARCH TOPICS

ABSTRACT

This research study examined the operations and performance of several facilities that were using the Lockheed DIALOG system for on-line searching of the ERIC data base. The study was done with the objective of identifying the factors that significantly influenced the productivity of terminal use. Detailed examinations were made of such aspects as the DIALOG system response time as a function of the time of day or day of the week; the search commands and logic used by each of the terminal installations for their operations; the mix of complex, medium or simple questions processed at each terminal location; and the extent and impact of the variant forms of descriptors in the file (e.g., singular and plural forms of the same term). Guidelines were prepared for the searchers to consider for pre-search and terminal activities. Timing studies were performed to suggest some terminal procedures that could increase average on-line search speeds.

FIGURES

Figure No.		Page
1.	Average Search Rates Realized for Each Terminal Installation	7
2.	Terminal Equipment Input Channel Alternatives Available For the Lockheed DIALOG Installations Studied	35
3.	Query Used for the DIALOG System Timing Exercise	46
4.	Search Times for All Searches Done by Each Searcher	49
5.	Cumulative Mean Search Time for Each Searcher	50
6.	Sequence of Search Times for Each Day of Timing Searches	51
7.	Range of Search Times Experienced With Timing Searches--Excluding Down-Time Searches (Maximum-Minimum Envelope)	53
8.	Range of Search Times Experienced With Timing Searches--Including Down-Time Searches (Maximum-Minimum Envelope)	54
9.	Average Search Time for All Timing Searches Started at Various Times of the Day (Excluding Down-Time Searches)	55
10.	Average Search Time for All Timing Searches Started at Various Times of the Day (Including Down-Time Searches)	56
11.	Average Search Time for All Timing Searches Run on a Given Day of the Week	57
12.	Illustration of Question and Search Definitions	61
13.	Command Utilization Per Question	69
14.	Command Utilization Per Search	73
15.	Mix of Question Complexity for Each Terminal	8[illegible]
16.	Relationship of Search Rates to Search Complexity	81
17.	Sample EXPAND Display	84-85
18.	Sample Page from Institutional Sources, Statistics and Postings	101
19.	Sample Page from Macmillan Descriptor and Identifier Usage Report	102
20.	Sample Page from ERIC Thesaurus	105

FIGURES (cont'd)

Figure No. | Page

21. Sample Page from Lockheed Term Frequency List 108
22. Sample Saved Search 110
23. Examples of Initialization Techniques 116-117
24. Sample EXPAND Display 118
25. Sample EXPAND Display For Related Terms 119
26. Example of an Incremental Search 127
27. Total Number of Citations Retrieved For the Incremental Searches 130
28. Percent of Total Citations Retrieved For the Incremental Searches 131
29. Percent of Relevant Citations Retrieved For the Incremental Searches 132
30. Example of Use of LIMIT/MAJ on Different Set Combinations 136
31. Example of RIE Issue Numbers Stored as an Inverted File 137

TABLES

Table No. Page

1. Average Search Rates Realized for Each Terminal Installation 8
2. Prior Evaluation Studies of On-Line Bibliographic Searching Systems 16
3. Terminal Equipment Used by the Installations Studied 37
4. Characteristics of Installations Studied 42
5. Questions, Searches, and Logical Operations Per Terminal 64
6. Command Utilization Per Terminal (By Number of Commands) 66
7. Command Utilization Per Terminal (By Terminal Time Used) 67
8. Command Utilization Per Question (By Number of Commands) 68
9. Command Utilization Per Question (By Terminal Time Used) 71
10. Command Utilization Per Search (By Number of Commands) 72
11. Command Utilization Per Search (By Terminal Time Used) 74
12. Average Command Execution Rate for Each Terminal Installation 75
13. Extent of Variant Forms Used or Not Used in 80 Real Searches 87
14. Characteristics of Variant Forms Used or Not Used in 80 Real Searches 88
15. Comparison of Original Search Results with the Result Obtained by Using All Possible Variant Forms of Descriptors and Identifiers 90
16. Functional Utilization of Commands (By Percent of Commands Used) 92
17. Terminal Productivity Related to Cost-Conscious Attitudes of Each Terminal Installation 95
18. Term Frequency Lists for the ERIC Data Base 107
19. Results of Timing Experiment for SELECTing 10 Descriptors 123
20. Incremental Effect of Adding One More Term to Each Facet 129

ACKNOWLEDGMENTS

Several people other than the authors contributed to the results of this study and deserve thanks and recognition. Don Thompson participated in the DIALOG system timing tests and did some of the initial literature review. Clare You performed the study of variant forms of ERIC descriptors and ran some of the searches to obtain data for the search guidelines. Allan Humphrey collected data from the site visits and was responsible for the collection of detailed data from the 15-day sample of terminal searches. Ken Renworth did most of the computer programming to generate the detailed terminal use statistics from the Lockheed data tapes. Pauline Atherton provided very helpful reviews and critiques at several points during the project. Clayton Perdue and Ron Heckart assisted in the preparation of this report.

We owe our thanks also to the staff members of the several ERIC searching sites that we visited during this study. They were very hospitable and made our work considerably easier and more pleasant.

And finally, we wish to acknowledge the help and assistance provided to us by the Lockheed staff members associated with this study, particularly Roger Summit and Mark Radwin. They were very responsive to our needs for data and other information relevant to this study.

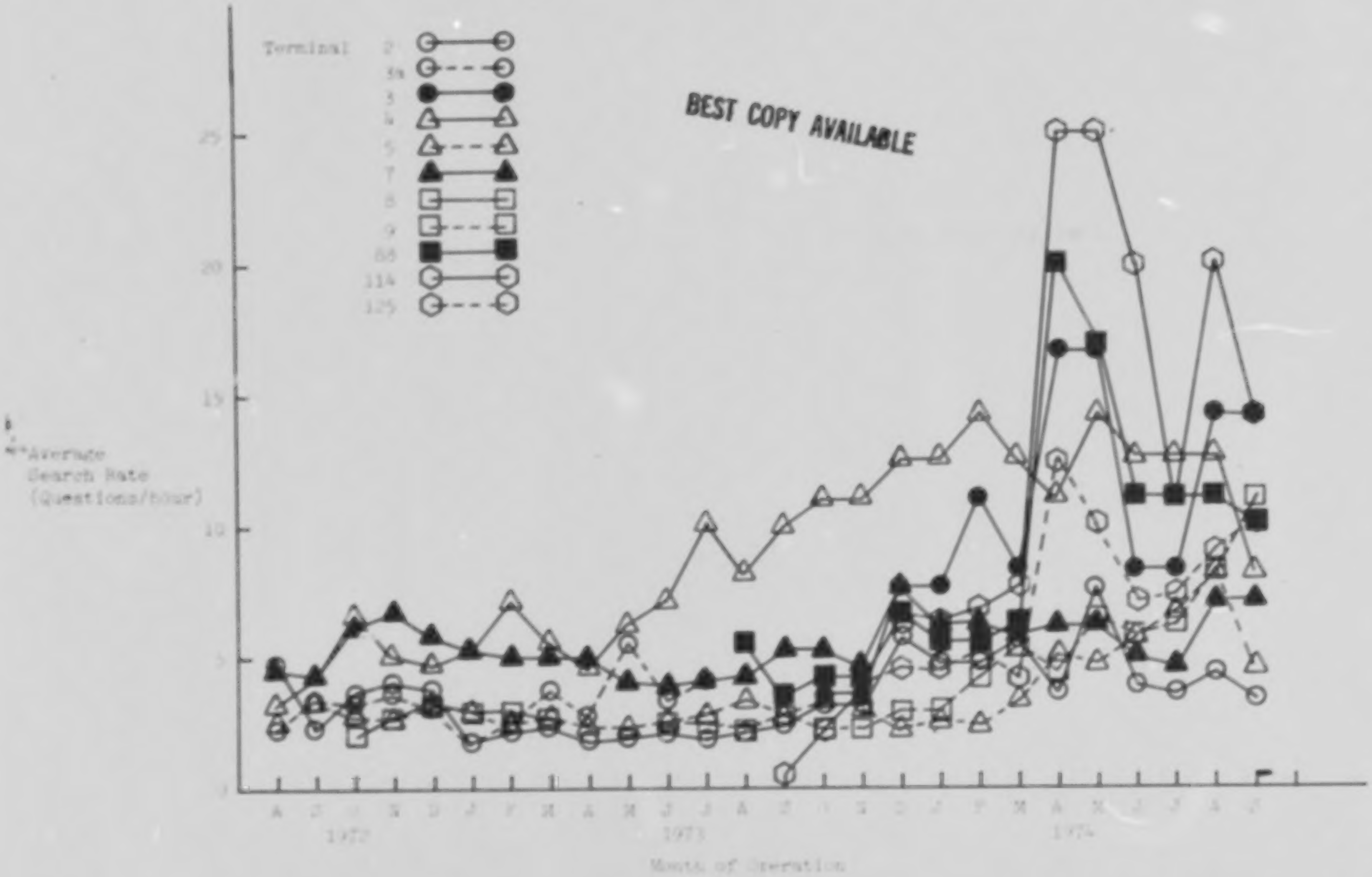
I. INTRODUCTION

A. THE PROBLEM

Since 1969 the Information Sciences Laboratory of the Lockheed Palo Alto Research Laboratory has offered on-line computer searching of the Educational Resources Information Center (ERIC) data base. During the past three years there has been a growing number of organizations conducting ERIC searches on-line using terminal equipment linked via telephone connections to the Lockheed computer facility in Palo Alto, California. Several installations have been heavy users of the service over a long period of time.

In studying the searching activity of the various ERIC service subscribers over a period of many months, it has been observed that wide variations exist among installations with respect to the average number of ERIC searches processed per unit of terminal time. Certain organizations consistently conduct more on-line searches per hour on their terminals than do others. Some installations consistently realize close to three times as many searches per hour as some other installations. This seems rather surprising considering that each installation searches the same data base, uses similar terminal equipment, receives similar instruction, uses similar support tools, and is served by the same central facility.

The extent of the variation in average search time among different organizations is shown in Figure 1, where the average search rate is given for 11 different terminals that actively used the ERIC/DIALOG system at some time during the period August 1972 through September 1974. (Not all of these organizations were still using ERIC/DIALOG during the Fall of 1973, the period examined most intensively during this investigation.) Some of these data were initially published in issues of the ERIC/DIALOG *Chronolog*, and are summarized in Table 1. It may be seen that while the average search rate fluctuates substantially from month to month, there are several installations that consistently process more questions per terminal hour than do several of the others. These installations, over a period of many months, have a typical average search rate of 9 or 10 questions per hour or more compared with 3 or 4 questions per hour for some of the other terminals. Note also that there has been a dramatic improvement in the search speeds for many of the installations since the early days of ERIC/DIALOG operation.



BEST COPY AVAILABLE

FIG. 1. Average Search Rate Realized for Each Terminal Installation

[illegible]

BEST COPY AVAILABLE

B. OBJECTIVES

The primary objective of this study was to identify and explain the reasons for the differences in average on-line search rates between different terminal installations. An additional objective was the development of a set of guidelines for conducting ERIC searches on-line via the Lockheed facilities that would enable the user to carry out the searching process in a more efficient and effective manner.

We were interested in identifying the factors that significantly influenced terminal productivity, with a view to establishing guidelines or suggested procedures that would permit the terminal operators to make the best use of their available facilities, and would provide some suggestions for additional system improvements.

Although the primary objectives were focused on searching the ERIC data base with the DIALOG system, it is expected that some of the findings will be applicable to other ERIC search systems and to other on-line search systems. With over 100 terminal installations presently searching the ERIC data base, and with more being added every day, this seemed to be a topic of increasing interest to a large number of organizations.

II. SUMMARY

The first objective of this study was to investigate various factors that could possibly influence the productivity of ERIC/DIALOG on-line searching and that might explain the wide difference in search productivity that was experienced by various terminal installations. Considering only the ERIC data base and the DIALOG search system, we have identified many different factors that could influence the rate at which searches were done at such terminals. Such factors include:

. computer and communication system loading (possibly reflected in different system speeds and response times for various hours of the day or days of the week)

. keyboard or typing skills of the searchers

. complexity of the questions being searched

. characteristics and speed of the terminal equipment used

. work habits and search formulation style of individual searchers

. extent of pre-search planning and work that is done before using the terminal

. availability and use of printed analyst reference tools

. extent of use of SEARCH SAVEs or other prior search efforts

. extent to which the terminal equipment is being used as an output device

. extent of use of operating shortcuts with the DIALOG system

. extent of relevant continuing education and association with other searchers

. subject expertise of the searcher and the installation

. cost-conscious attitude of the searcher and the installation

. degree of user versus intermediary searching, and extent of user involvement in the on-line interaction

Data was collected during this study that focused on several of these factors, however we do not really know the impact or influence of all of these factors.

We have come to the conclusion that there is no single dominant factor that strongly influences the search speeds. We see instead a complex pattern of many influences at work. At this time we cannot even rank all of these factors in terms of their relative influence on search speeds, and feel confident about that ranking. However we can sift out and identify some of the most important, and some of the least important factors.

The factors that seemed to have little influence on search speeds were:

- computer system loading (and time of day or day of the week). Timing tests showed very little difference in search speed for various times of the day, or days of the week.
- keyboard or typing skills of the searcher

Factors that seemed to have the most influence on search speeds were:

- work habits and search formulation styles (e.g. question complexity, recall/precision goals) of the searchers
- extent of pre-search preparation for each question
- cost-conscious attitude of searchers
- searcher familiarity with the data base and operating skill with the on-line system (i.e. how good a "driver," and how many shortcuts does the searcher know and use?)
- characteristics and speed of the terminal equipment

Unfortunately only the last of these factors is supported by any solid evidence from this study. The other factors are on the list primarily on the basis of observation, personal experience, and discussions with other searchers; however we are confident in their selection. The remaining factors on the initial list have some influence, but to an extent yet to be determined.

The second objective of this study was to develop some guidelines to help improve the performance of ERIC/DIALOG searching. As a result of some timing exercises and other controlled experiments, and discussions with searchers at other installations, specific performance-improving practices were suggested with regard to such aspects as:

- deciding when to do a search manually instead of on-line
- handling variant forms of subject terms
- use of the various analyst support tools
- use of the SEARCH SAVE feature

- ways to limit search output
- initialization procedures
- ways to SELECT Descriptors
- deciding how many terms to use to adequately describe a search topic.

III. REVIEW OF EXISTING LITERATURE

A literature search was undertaken to determine whether evaluation methodologies had been suggested or used which could be applied to our study of the productivity of on-line terminals accessing the ERIC data base through Lockheed's DIALOG. Relevant citations are given at the end of this section.

A. IMPORTANT MONOGRAPHS

Anyone beginning an evaluation of an information retrieval system should be familiar with the following landmark monographs in the field.

Lancaster's Information Retrieval Systems. Characteristics, Testing and Evaluation (1968), is an outstanding book, which received the American Society for Information Science's award for best information sciences book of the year in 1970. It is concerned primarily with intellectual factors that significantly affect the performance of information retrieval systems (batch or on-line): indexing policy and practice; vocabulary control; searching strategies; interaction between the system and its users. Recall and precision are selected as the most important measures of system performance.

Chapter 2 of King and Bryant's The Evaluation of Information Services and Products discusses kinds of measures that can be used to evaluate information retrieval systems, including recall, precision, fallout and generality ratios, and total retrieval. Measures and techniques suitable for "macroevaluation" (gross measures of input, output and effectiveness, made at minimum cost, suitable for support of decisions affecting funding and administration), and "microevaluation" (identification and diagnosis of failing components of information systems) are presented in Chapters 3 and 4.

Lancaster's Information Retrieval On-Line (1973), particularly Chapters 8 and 9, describes a number of systems currently in operation, discusses evaluation methodologies, and summarizes several evaluative studies done to date. Several chapters contain extensive bibliographies.

B. BIBLIOGRAPHIES AND REVIEWS

There is a great deal of literature about on-line information retrieval systems. Much of it is devoted to descriptions of specific systems and their features. No attempt has been made to include this literature in the present compilation, since it is included in several published bibliographies. There is also a quantity of literature on the evaluation of information retrieval systems (not necessarily on-line systems). Some general papers have been included in addition to the evaluation studies listed in the next section.

In recent years, the user interface has been the subject of considerable attention. John Bennet's chapter of the ASIS Annual Review of Information Science and Technology for 1972 deals comprehensively with "The User Interface in Interactive Systems". Thomas Martin's chapter for the 1973 ASIS Annual Review on the same subject, focuses primarily on the conceptual aspects of interaction. The 1971 AFIPS workshop proceedings, Interactive Bibliographic Search: the User-Computer Interface, edited by Don Walker, contains an extensive classified bibliography.

A bibliography on Evaluation of Document Retrieval Systems covering literature published up to early 1968 was published by Wiederkehr. and Beth Krevitt published a bibliography on Evaluation of Information Systems that covered the 1967-1972 literature.

C. PRIOR EVALUATION STUDIES

Wiederkehr (pg. 14) discussed the appropriateness of measures of effectiveness, and of efficiency as follows:

> "An appropriate measure to be used as a criterion for evaluating an information retrieval system should account for both how effectively the objectives are being met as well as how efficiently resources are being used. Consequently, it is desirable to have measures of effectiveness, such as how many useful documents were retrieved, and measures of efficiency, such as the cost and time. Recall and precision only partly satisfy this desire.
>
> In the research and development phase of any system, the primary objective is to demonstrate the technical feasibility of the system. Accordingly, effectiveness is of prime importance and efficiency is often ignored. Once the technical feasibility of the system has been proven, the objective shifts to demonstrating the economic feasibility of the system. In most operating systems economic feasibility is of prime importance, in which case both the effectiveness and the efficiency should be taken into account.
>
> Since most efforts to date concerning the evaluation of information retrieval systems have treated systems in the research and development phase, most of the measures considered have been measures of effectiveness, such as recall and precision. However, as the systems become operational on a large scale, measures of efficiency and overall measures which account for both effectiveness and efficiency are anticipated."

Our study was concerned primarily with measures of productivity and efficiency.

A number of evaluations of on-line systems have been carried out. Most of these focused on measures such as recall and precision; most dealt with effectiveness of search formulation and with human factors such as training required, attitudes, or frustration. Few have dealt with efficiency or productivity as a primary factor: effectiveness has been much more the focus.

Lancaster suggests that there are two basic ways of collecting evaluative material, which may be used separately or in combination: through a series of survey forms (e.g., pre- and post-retrieval questionnaires), or through the terminal itself -- using computer data collection techniques. (Lancaster, 1973, p. 157.)

Table 2 summarizes the aspects covered, and the techniques used, in prior evaluation studies of on-line bibliographic searching systems.

TABLE 2

PRIOR EVALUATION STUDIES OF ON-LINE BIBLIOGRAPHIC SEARCHING SYSTEMS

	NATURE OF STUDY		ASPECTS COVERED						TECHNIQUES USED				
INVESTIGATOR	Theoretical	Experimental	Recall/ Relevance	Query Language & Formulation	Search Failure Analysis	Human Factors	Time & Cost	Other	User Survey	Examination of Search Output	Audit of Computer Records	Comparative Searches	Other
Atherton (SUPARS)		X	X	X	'HELP' analysis	X	X		X	X	X		
Back, 1971	suggested research			X		X					X		tape recording
Coles (NASA/RECON)		X	X				X		X	screening			
Cook (SUPARS)		X	X	X	'HELP' analysis	X	X		X	X	X	X	
Katter	X					X			X		X		
King (CIRCOL)		X	X	X		X		indexing	X	X		X	
Lancaster (MEDLARS) (1968, 1969)		X	X	X	X	X	response time	indexing, novelty, serendipity	X	X		X	
Lancaster (AIM-TWX)		X	X	X	X	X			X	X		X	
Lancaster (EARS) (1972)		X	X	X	X	X			X	X		X	
Meister (NASA/RECON)		X		X		X	X		X			X	
Melnyk		X				X			X				observation through glass partitions
Mittman (RIQS)		X		X		X	X				X		
Martin	comparative analysis			X				system features					
Rosenberg		X				X			by tape recording		X		tape recording:user "thinks aloud"
Summit (RECON)		X		X		X		system features			X		
Summit (DIALOG)		X		X		X		system features	X				
Timbie (DIALOG:ERIC)		case studies		X		X	X	system features	X				
Treu		X		X		X	X	hardware features			X		"Transparent stimulation"

D. OTHER STUDIES

Several of the studies included in Table 2 devote considerable attention to specific system features. Other papers that discuss system design features in more general terms are those by Back (1972), Bennett (1971), and Martin (1974). Martin's paper compares specific system features of 11 on-line searching systems. The report is aimed at system designers, and is not intended to be a system selection guide. This report is the third in a sequence which began with the AFIPS workshop (Walker) and continued with the AFIPS/ASIS workshop (Martin, JASIS 1973). Fife's review paper discusses over 50 technical features of some 46 interactive information systems. This paper is designed to help with state-of-the-art assessments prior to system selection. System features are also discussed in Interactive Bibliographic Systems, where papers and discussions from a 1971 conference are presented.

Cooper has proposed a measure of effectiveness involving a weighted output and the number of citations the user desired. Tell and Williams discuss the inverse weighting (value) of index terms according to their frequency of use (e.g., specific terms with few postings are associated with high weights; terms with many postings, with low weights).

A few articles dealing specifically with ERIC though not necessarily in an on-line context, have been included. Fry and Tell mention the quality of RIE material; Jewell discusses search strategies specifically for the ERIC data base.

Although Mittman, Treu, the SUPARS group, and others have kept logs of terminal activity, nothing was found in this literature search which was directly applicable to our primary purpose: to investigate the productivity, or factors affecting the productivity, of several different terminals accessing a central on-line data base.

Annotated citations for the major papers identified in this literature search are given in the next section.

E. OTHER WORK IN PROGRESS

Several studies are currently in progress at other locations that may provide additional findings that are relevant to this effort. Dave Penniman of the Battelle Memorial Institute is presently completing his dissertation at Ohio State University on a topic related to on-line bibliographic searching systems. As part of that study he has performed a detailed statistical study of the terminal transactions for Battelle's BASIS system. The analysis included time distributions of 11 major functions performed at the terminal (e.g., initialize, search index, formulate logic, print) for each terminal and each of several major data bases. This work was reported at the 1974 ASIS annual meeting but has not been published yet.

The BioSciences Information Service (BIOSIS) has been running an experimental on-line search service on the BIOSIS tapes, using the STAIRS software in conjunction with 26 terminal installations of the SUNY Biomedical Communications Network. As part of this experiment, a series of 20 test search statements was given to each installation for a controlled test of searching performance. Because the same 20 questions were given to each installation, it will be possible to analyze the different approach taken by 26 different analysts to code and run the same information requested. This should provide useful indications of the variability of analysts' approaches to the same problem. The experiment was discussed briefly at the 1974 ASIS meeting by Kay Durkin and is discussed in the conference proceedings. However, the full report of the research results is not expected to be available until sometime in 1975.

F. REFERENCES

Atherton, Pauline A., K. H. Cook and J. Katzer. *Free Text Retrieval Evaluation*. Syracuse, N.Y.: Syracuse University, School of Library Science, 1972.

An extensive series of SUPARS experiments was conducted in 1971-72. Reaction to the system was evaluated by 63 telephone interviews administered to random samples of users and non-users, a semantic differential (described in the 1972 Katzer article) and an analysis of requests for help made through a telephone aid service which was available to searchers at the terminal.

Ba[illegible], Harry B. "What Information Dissemination Studies Imply Concerning the Design of On-Line Reference Retrieval Systems," *Journal of the American Society for Information Science* 23:3 (May-June 1972) 156-163.

For a computer-based system to be accepted and used, it must be designed so that the effort required to obtain pertinent references from the computer is not much greater than the effort required using other methods.

The characteristics of informal methods of information gathering suggest five ways for minimizing the human effort expended in retrieving references from an on-line system.

1. Allow the user to shape the interaction to fit his needs.
2. Retrieve few irrelevant references.
3. Furnish references to the appropriate type of document (e.g., theoretical discourse, description of an application, review article, etc.).
4. Provide direction for further search.
5. Deliver screened and evaluated references.

Back, Harry B. and Richard L. Van Horn. "A System to Improve the Availability and Usefulness of Management Science Knowledge," in Donald E. Walker, ed., *Interactive Bibliographic Search: The User/Computer Interface*. Montvale, N.J.: AFIPS Press, 1971, 19-43.

The user interface features of a prototype retrieval system are described. A research experiment is suggested, results of which could be used to successively modify the interface design. In this research, users would be given standard retrieval tasks. User "protocols" would be monitored. Users would be asked to "think aloud" and would be monitored by tape recorder in addition to computer terminal records. Both written and oral user actions in problem solving would be analyzed.

Bennett, John L. "Interactive Bibliographic Search as a Challenge to Interface Design," in Donald E. Walker, ed., Interactive Bibliographic Search: The User/ Computer Interface. Montvale, N.J.: AFIPS Press, 1971, 1-18.

This paper was the challenge paper for the AFIPS Workshop on "The User Interface for Interactive Search of Bibliographic Data Bases."

Challenge points were:

(A) Characteristics of the searchers served by the facility.
(B) Conceptual framework presented to the searcher.
(C) Role of feedback to the searcher during search.
(D) Operational characteristics of the facility: the command language, display of formats, response time
(E) Constraints of the terminal and techniques to ameliorate them.
(F) Effect of the bibliographic data base on the user interface for search.
(G) Introducing the search facility to the user.
(H) Role of evaluation and feedback in the redesign cycle.

Bennett, John L. "The User Interface in Interactive Systems," in Carlos A. Cuadra, ed., Annual Review of Information Science and Technology. Vol. 7. Washington, D.C.: American Society for Information Science, 1972, 159-196.

A comprehensive review with 115 references.

Coles, Victor L. "Remote Evaluation of a Remote-Console Information-Retrieval System (NASA/RECON)," in Interactive Bibliographic Systems. Proceedings of a Forum Held at Gaithersburg, Maryland, October 4-5, 1971. Oak Ridge, Tenn.: U.S. Atomic Energy Commission, Office of Information Services, April 1973, 133-142. Open discussion, 142-149. CONF-711010.

Continuing evaluation is sought at the NASA Scientific and Technical Information Office. NASA/RECON search results are sent to users accompanied by an evaluation form. The search procedure features delegated searching via written requests. Search analysts screen the printed output before sending it to the user.

Cook, K. H., L. H. Trump, P. Atherton and J. Katzer. Large Scale Information Processing Systems. Final Report. Syracuse, N.Y.: Syracuse University, School of Library Science, 1971. 6 vols.

The full report on the SUPARS experiments.

Cooper, W. S. "Expected Search Length: A Single Measure of Retrieval Effectiveness Based on the Weak Ordering Action of Retrieval Systems," American Documentation 19 (January 1968) 30-41.

Given a set of retrieved documents ordered by expected relevance, a measure of effectiveness is obtained relative to the user's quantification of the number (n) of relevant documents desired. The expected search length is defined to be the number of nonrelevant documents preceding the nth relevant document. This can be compared against an expected search length of a hypothetical, randomly ordered system output. The fractional reduction in expected search length in going from the random to the actual system is called the mean expected search length factor.

Fife, Dennis W., and others. A Technical Index of Interactive Information Systems. Final Report. Washington, D.C.: National Bureau of Standards, March 1974. 79p. FGK56375. ED-092 163.

The technical features and operational status of interactive information systems, i.e. those providing a conversational usage mode to a non-programer through a data terminal device, are reviewed. The review is designed to aid information specialists in the state-of-the-art assessments preparatory to a detailed system selection procedure. It contains an index: 46 systems are listed by trade name. The index provides information about over 50 technical features. Information is based primarily on documentation received during 1972 and 1973. In addition, there are aids and examples contributing to the intended use of the index.

Fry, Bernard M. Evaluation Study of ERIC Products and Services. Summary Volume. Final Report. Bloomington, Ind.: Indiana University, Graduate Library School, March 1972. 51 p. ED-060 922.

Although the scope of this evaluation specifically excludes evaluation of the ERIC tape data bases, it is of interest for some of the comments about the data in RIE. Data gathered from individual users' responses, site interviews, and advisory panels suggested the following changes or improvements in RIE should be studied: (partial list)

- merging institutional entries without regard to sub-divisions
- coding level (age, elementary, high school, etc.)
- coding type (speech, survey, report, etc.)
- omitting or flagging non-available documents
- indexing consistency as between general or specific
- correcting unevenness in quality of documents.

RIE was evaluated high on its range of topics, the contents of resumes, and the indexing system, but relatively low in other characteristics, including quality of material selected and timeliness.

Interactive Bibliographic Systems. Proceedings of a Forum Held at Gaithersburg, Maryland, October 4-5, 1971. Oak Ridge, Tenn.: U.S. Atomic Energy Commission, Office of Information Services, April 1973. 205 p. CONF-711010.

This volume contains papers and discussions in the areas of user interface, system configuration, economics and performance, and future developments. Many of the papers give detailed operating experiences. The open discussions are particularly interesting; they include timing and cost information. Van Wente quotes an average response time for a RECON command as 40 seconds (average range including BEGIN, SELECT, COMBINE and other commands). Average sit-down time for a RECON user at Goddard Space Flight Center, NASA, was quoted as one-half hour, with thorough and extensive searches such as those needed before starting a new research project, easily lasting an hour. (p. 16, 17.) Coles reports slightly longer search times (45 minutes), though overall times for delegated searching average 1-1/2 hours.

Jewell, Sharon and W. T. Brandhorst. *Search Strategy Tutorial; Searcher's Kit*. Washington, D.C.: National Inst. of Education, October 1973. 86 p. ED-082 763.

From the ERIC Data Base Users Conference, Columbus, Ohio, October 10-12, 1973. This document is the workshop manual used in a three-hour tutorial session on search strategies. The discussion of the input phase of a computer search covers identification of the user population, receiving the inquiry, and the types of services offered. General principles of good searching, search theory and general manipulative capabilities are discussed, as well as specific properties of the ERIC system that affect computer search capabilities. A practice session is included. The output phase of a computer search includes a discussion of output formats, output evaluation, and statistical records-keeping.

Katter, Robert V. "Insights in Implementing the Redesign Cycle," in *Interactive Bibliographic Systems*. Proceedings of a Forum Held at Gaithersburg, Maryland, October 4-5, 1971. Oak Ridge, Tenn.: U.S. Atomic Energy Commission, Office of Information Services, April 1973, 175-182. CONF-711010.

Four classes of feedback are discussed:

(1) System contact and use statistics.
(2) User commentaries.
(3) Output-efficiency evaluations.
(4) Interaction recordings.

"Sampling techniques to record terminal interactions and separate off-line programs to analyze the data can be combined to provide efficient recording and reduction of data."

Katzer, Jeffrey. "The Cost-Performance of an On-Line, Free Text Bibliographic Retrieval System," Information Storage and Retrieval 9 (1973) 321-329.

"Performance measures, such as recall and precision, do not supply any information about the operating efficiency of a system. What is needed, and what has been suggested for some time now, is performance characteristics paired with cost measures."

An estimate of recall was tabulated against costs at 10-90% estimated recall levels. Results indicated that, for all but the lowest recall levels, SUPARS I had better cost-performance operating characteristics under the restrictions of simple logical operators (using only OR and AND logic, versus using OR, AND, NOT logic combined with word root searching). Another finding was that SUPARS I was very expensive.

SUPARS II findings indicated that on-demand access to the index or dictionary contributes significantly to improving the cost performance.

Katzer, Jeffrey. "The Development of a Semantic Differential to Assess Users' Attitudes Toward an On-Line Interactive Reference Retrieval System," Journal of the American Society for Information Science 23:2 (March-April 1972) 122-128.

A user questionnaire employing nineteen 7-interval adjective scales, such as fast-slow, active-passive, good-bad is described.

The major finding of the study was that users reliably respond toward such a system. Their affective responses can be conceptualized into three independent components: (1) the evaluation of the system; (2) the desirability of the system; and (3) the enormity of the system.

King, Donald W. and Edward C. Bryant. The Evaluation of Information Services and Products. Washington, D.C.: Information Resources Press, 1971. 306 p.

Chapter 2 contains a good discussion of performance measures based on user relevance judgments, including recall, precision, fallout and generality ratios. Macroevaluation and microevaluation are presented in Chapters 3 and 4.

King, Donald W., et al. Comparative Evaluation of the Retrieval Effectiveness of Descriptor and Free-Text Search Systems Using CIRCOL (Central Information Reference and Control On-Line). Rockville, Md.: Westat Research, Inc., January 1972. RADC-TR-71-311. ED-068 137. Available through NTIS (AD-738 299).

The study compares the retrieval effectiveness of two alternative input and search systems in terms of such measures as recall, fallout, precision, and total retrieval. One system operates using manually indexed document files searched by controlled vocabulary, while the other employs full text input using natural language searching. The results indicate that the two systems perform at approximately the same level of effectiveness, although estimated average total retrieval was found to be slightly greater for free-text searching than for descriptor searching at all levels of recall.

Krevitt, Beth and Belver Griffith. "The Evaluation of Information Systems: A Bibliography 1967-1972," in Information Pt. II, vol. 2, no. 6. New York: Science Associates International, 1973. p. 1-34.

The scope of this classified bibliography is limited to the design, testing and evaluation of information storage and retrieval systems. Contains sections on evaluation techniques and on on-line interactive systems.

Lancaster, F. Wilfrid. Evaluation of the MEDLARS Demand Search Service. Washington, D.C.: National Library of Medicine, January 1968. 276 p. Available from NTIS (PB-178 660).

This report presents the results of a detailed analysis by the National Library of Medicine of the performance of MEDLARS in relation to 300 actual requests made to the system in 1966 and 1967. Delegated searches (demand search bibliographies) were requested in person, by mail directly, or by mail through a librarian or information specialist. A MEDLARS search was performed, and in addition to the printed output, photocopies of 25 to 30 retrieved articles (selected by random sampling if total retrieval was larger than 30) were sent to the user, who evaluated each article on a three point scale (major, minor, no value), as well as on a fourth point ("glad to learn of article's existence because of some other need or project"). A recall base was obtained from known relevant articles supplied by the user, supplemented by a manual literature search. Recall, precision, and "novelty" ratios were obtained. (The novelty ratio is based on the "other need" answer.) The system was shown to be operating, on the average, at about 58% recall and 50% precision. However, search results were widely scattered; some achieved high recall and high precision; others achieved completely unsatisfactory recall results. A detailed failure analysis was performed.

Lancaster, F. Wilfrid. Information Retrieval Systems. Characteristics, Testing and Evaluation. New York: Wiley, 1968. 222 p.

This book received the American Society for Information Science's award for best book of the year on information science, in 1970. It is concerned primarily with "intellectual" factors that significantly affect the performance of all information retrieval systems, namely: indexing policy and practice; vocabulary control; searching strategies; interaction between the system and its users.

Recall and precision are selected as the most important measures of system performance; indexing, search strategies and other factors influencing this performance are discussed, using detailed examples. Failure analysis plays an important part in the overall analysis.

Lancaster, F. Wilfrid. "MEDLARS: Report on the Evaluation of its Operating Efficiency," American Documentation 20 (1969) 119-143.

A comprehensive program to evaluate the performance of MEDLARS was conducted by the National Library of Medicine in 1966 and 1967. This report describes the methodology used and presents a summary of the principal results, conclusions, and recommendations. The detailed report on this study is listed above (PB-178 660).

Lancaster, F. Wilfrid. Evaluation of On-Line Searching in MEDLARS (AIM-TWX) by Biomedical Practitioners. Urbana, Ill.: Illinois University, Graduate School of Library Science (Occasional Papers 101), February 1972. 21 p. ED-062 989.

The purpose of the investigation was to determine how effectively biomedical practitioners, with a minimum of introduction to the system, can conduct on-line searches to satisfy their own information needs. Forty-eight searches were conducted by biomedical practitioners on Abridged Index Medicus (AIM-TWX). Trained search analysts then structured and conducted searches on the same subject. It is concluded that many biomedical practitioners could exploit AIM-TWX profitably with the minimum of introduction to the system and without the necessity of using a trained MEDLARS analyst. Limitations of the ELHILL search system (SDC's ORBIT as modified for National Library of Medicine use) mentioned were: ELHILL should be less error-sensitive; more cross references are needed in the vocabulary file. Potential improvements suggested include term weighting; visual displays (CRT type); clustering techniques whereby documents "like" a given document could be found; and acceptance of approximate keywords.

Lancaster, F. Wilfrid and E. G. Fayen. Information Retrieval On-Line. Los Angeles, Calif.: Melville, 1973. 597 p. (A Wiley-Becker & Hayes Series Book.)

This book provides a broad survey of the characteristics, capabilities, and limitations of present on-line interactive systems for bibliographic search and retrieval. The emphasis is on the design, evaluation and use of such systems, primarily from the viewpoint of the planner and manager of information services. It is oriented toward the "intellectual" aspects of information retrieval rather than the hardware or programming aspects. Chapter 8: Evaluating Effectiveness of the System, and Chapter 9: Operating Experience and Evaluation Results, discuss evaluation methodologies and results from the several on-line search systems, including AIM-TWX, DIALOG, SUPARS and others.

Lancaster, F. W., Richard L. Rapport and J. Kiffin Penry. "Evaluating the Effectiveness of an On-Line, Natural Language Retrieval System," *Information Storage and Retrieval* 8:5 (October 1972) 223-245.

An evaluation of the Epilepsy Abstracts Retrieval System (EARS) was performed. Searches were conducted on the on-line system and evaluated in terms of recall, precision, and general user satisfaction. Searchers (who were doctors, not search analysts) filled out forms before starting to search, including identification, time started, relevant abstracts retrieved (these were to be looked up by number in the hard copy of Epilepsy Abstracts, located close to the terminal), and total elapsed time. Parallel searches were conducted by experienced searchers on the same topics.

Martin, Thomas H. *A Feature Analysis of Interactive Retrieval Systems*. Stanford, Calif.: Stanford University, Institute for Communication Research, September 1974. 100 p. SU-COMM-ICR-74-1. Available from NTIS.

The command language features of eleven different on-line information retrieval systems are presented in terms of the functional needs of a searcher sitting at a terminal. Functional areas considered are: becoming familiar with the system, receiving help when in trouble, regulating usage, selecting a data base, formulating simple queries, expressing single concepts, interconnecting concepts, displaying results simply, and controlling the display. Features felt most essential to on-line searching are live help, users' guides, boolean operators, search field control, suffix removal, relational operators, dictionary access, request sets, search review, predefined formats, on-line formatting, and off-line printing. It is concluded that no sharp distinction exists between management information and bibliographic retrieval. The report is intended for use by designers of interactive retrieval systems and by students of system design.

Martin, Thomas H. "The User Interface in Interactive Systems," in Carlos A. Cuadra, ed., *Annual Review of Information Science and Technology*. Vol. 8. Washington, D.C.: American Society for Information Science, 1973, 203-219.

This review focuses in the conceptual aspects of interaction. 48 references.

Martin, Thomas H., James Carlisle and Siegfried Treu. "The User Interface for Interactive Bibliographic Searching: An Analysis of The Attitudes of Nineteen Information Scientists." Journal of the American Society for Information Science 24:2 (March-April 1973) 142-147.

Results of a questionnaire administered to 19 information scientists at an AFIPS/ASIS sponsored workshop on user computer interface. This workshop was a sequel to the AFIPS workshop reported by Walker. 147 propositions (software and hardware features or response patterns) were presented and respondants were asked to rate them on a five point scale from "too rigid" to "too flexible." Cost was not to be considered. Consensus was reached on 70 items at the .025 probability level.

Meister, David and Dennis J. Sullivan. Evaluation of User Reactions to a Prototype Information Retrieval System. Canoga Park, Calif.: Bunker-Ramo Corp., October 1967. 62 p. (NASA-CR-918). ED-019 094.

This early evaluation of the experimental RECON retrieval system as implemented by Bunker-Ramo Corp., was conducted using two separate measures to determine acceptability and usability: (1) frequency of system usage, and (2) personal opinion of the user population. A second method of evaluation consisted of measuring the accuracy and speed of RECON as compared with the major existing information retrieval method, an off-line computer search, formulated by librarians.

Users were satisfied that on-line searching was faster than off-line batch searching or manual searching, but felt that RECON's response time was very slow.

Melnyk, Vera. "Man-Machine Interface: Frustration." Journal of the American Society for Information Science 23:6 (November-December 1972) 392-401.

As an exploration of the frustration experienced by users of an on-line interactive retrieval system, students participated in an experiment using an experimental reference retrieval system for library literature. Subjects were monitored by observation through glass partitions, in addition to a questionnaire on their emotional state. The control group received instruction and a plan for searching; the experimental group received a demonstration on one system, but had to use two other, undemonstrated systems. The experimental group experienced much more frustration.

Mittman, Benjamin and Wayne D. Dominick. "Developing Monitoring Techniques for an On-Line Information Retrieval System," Information Storage and Retrieval 9 (1973) 297-307.

Northwestern University's RIQS (Remote Information Query System) generates the RIQSLOG, which logs all user actions and system actions.

This log is processable by the RIQS programs. The monitor was designed to provide data for analysis of system response, user errors, query complexity, use of resources such as central processor, time required for searching, etc. During winter quarter 1972, students were monitored in some 130 on-line sessions, containing approximately 625 individual queries against a data base which contained 157 records from articles which had been published in the Communications and Journal of the Association for Computing Machinery.

Average real time per session (probably analogous to DIALOG's search) was 23.9 minutes, with a range from 0.5 to 111.3 minutes. CPU time per query averaged 0.6 seconds, and ranged from 0.04 to 28.9 seconds. Real time per query (probably analogous to DIALOG's question) broke down as follows:

- in 72 percent of the queries, real time for query input was less than 3 minutes
- in 92 percent of the queries, real time for query input was less than 6 minutes
- in 4 percent of the queries, real time for query input was greater than 8 minutes, with a maximum of 22 minutes.

A ratio was plotted: real time for query input over total real time for input and execution of that query. The ratio lay between 0.8 and 1.0, indicating that a very substantial amount of the real time for performing a search is attributable to entering the query into the system.

The initial attempt at relating query complexity to search time was to use a simple count of search terms. Large numbers of search terms associated with relatively little CPU time were observed for queries which tended to use AND operators (causing the algorithm to terminate execution of the Boolean combinations at the first false comparison).

A performance equation was developed which involved CPU predicted time, number of words generated to output reports, number of records scanned, and a measure of statement complexity. The measure of statement complexity "could not be obtained deterministically," but was manually assigned from a visual scan of the text. (The article does not elaborate further on the measure of statement complexity.)

Rosenberg, Victor. "A Technique for Monitoring User Behavior at the Computer Terminal Interface," Journal of the American Society for Information Science 24:1 (January-February 1973). 71.

Description of a "two track" user behavior observation. Track one is a printout of all communications between user and computer, with a time log in the margin indicating elapsed time. Track two is obtained with a tape recorder. The user is asked to "think aloud," to state the problem, the developing search strategy, evaluation of system performance, etc. This is transcribed and time-keyed, as was track one and the two are compared. Rosenberg comments on the difficulty of dealing with the resulting mass of data but suggests that it can be worth it for valuable insights obtained.

Summit, Roger K. "DIALOG and the User -- an Evaluation of the User Interface with a Major On-Line Retrieval System," in Donald E. Walker, ed., *Interactive Bibliographic Search: the User-Computer Interface*. Montvale, N.J.: AFIPS Press, 1971, 83-94.

A description of the DIALOG retrieval system is given. Ease of use is stressed. Additional details on the evaluations are described in Timbie.

Summit, Roger K. *Remote Information Retrieval Facility*. Palo Alto, Calif.: Lockheed Missiles and Space Co., April 1969. 44 p. (NASA-CR-1318).

Describes the use of DIALOG by NASA. It was found that end-users tended to use more complex logic and more terms than intermediary searchers. Complexity was indicated by the number of logical AND connectors used by searches.

Tell, Bjorn V., and others. *The Use of ERIC Tapes in Scandinavia; Searching with Thesaurus Terms in Natural Language*. Strasbourg, France: Council of Europe; and Stockholm, Sweden: Council for Cultural Co-operation; Royal Institute of Technology, 11 November 1972. 23 p. (ECS-DCC-72-15). ED-072 794.

This is a description of the batch processing SDI system used in Sweden, using the ABACUS and VIRA programs to search the ERIC files. The high noise level of the ERIC data base is mentioned; in one case (a search on audiovisual aids for the mentally retarded) it was found to be about 40%. This is considered quite high, considering that ERIC is a central data base for this sort of request.

A weighting procedure is suggested, based on term-usage frequency. Tell suggests that high frequency terms are looked upon as having less value than those with low frequencies. (Williams mentions this also.) A value assignment of 1/n, where n is the number of term postings over a large sample from each data base, say 30,000 references, would allow output printout to be ordered according to the sum of specificity of retrieval terms.

(Note: Simply because a term has low "information value" does not mean that it is useless in a search. It may provide a background against which other terms must be searched.)

Timbie, Michele and Don Coombs. An Interactive Information Retrieval System; Case Studies on the Use of DIALOG to Search the ERIC Document File. Stanford, Calif.: Stanford University, ERIC Clearinghouse of Educational Media and Technology, December 1969. 90 p. ED-034 431.

User studies. A synopsis of this study is reported in Lancaster (Information Retrieval On-Line, 1973).

Treu, Siegfried. A Computer Terminal Network for Transparent Stimulation of the User of an On-Line Retrieval System. Washington, D.C.: National Bureau of Standards, Center for Computer Sciences and Technology, July 1972. 39 p. (NBS-TN-732). ED-070 461.

A computer terminal network to enable "transparent stimulation" of the user of an on-line retrieval system has been designed, implemented, and pilot tested. Its basic purpose is to provide a suitable and effective framework and methodology for experimental identification/validation of those human characteristics which should be recognized/reinforced in man-computer interface design. The rationale behind the transparent stimulation approach is presented and the methodology employed for such real-time, unobtrusive scanning and manipulation of the man-computer dialogue is described. A general overview of the hardware and software features of the implemented stimulation network is included.

Treu, Siegfried. "A Conceptual Framework for the Searcher-System Interface," in Donald E. Walker, ed., Interactive Bibliographic Search: The User-Computer Interface. Montvale, N.J.: AFIPS Press, 1971, 53-66.

This article discusses the terminal dialogue monitoring capability being developed at the National Bureau of Standards. Transparent stimulation (whereby a person at a remote location causes the computer to prompt attitude-related questions) is also discussed.

Treu, Siegfried. "Techniques and Tools for Improving the Interactive System Interface," in Interactive Bibliographic Systems. Proceedings of a Forum Held at Gaithersburg, Md., October 4-5, 1971. Oak Ridge, Tenn.: U.S. Atomic Energy Commission, Office of Information Services, April 1973, 32-38.

Two specific data collection tools have been developed at the Center for Computer Sciences and Technology of the National Bureau of Standards. The dialogue monitor records the time of sending and receipt of each command in both directions. Thus data on the system response time after the user hits the carriage return key, the time the system takes to transmit an entire message, or the time from receipt of the system message to the next user input (i.e., user think time and keying time), may be obtained. Transparent stimulation is recommended as an unobtrusive monitoring technique, and is described as follows:

> "The software is loaded by the (remote) observer, who has both a teletypewriter and a specially constructed emotion-reason indicator device available to him. Whenever he wishes (or the software demands) during the course of a user-system dialogue, the observer can request that the user indicate his current level of satisfaction (e.g., whether annoyed, frustrated, happy as well as the reason therefore. The specially designed and constructed terminal that is available to the user enables him to push the appropriate labeled buttons after being prompted by a light and buzzer. The observer has an exact copy of that terminal except that it has lights where the user terminal has buttons and a prompt button in the place of a light."

The messages created in the transparent stimulation mode are recorded by the dialogue monitor, along with their times.

Walker, Donald E., ed. Interactive Bibliographic Search: The User-Computer Interface. Proceedings of a Workshop held in Palo Alto, California, on 14-15 January 1971. Montvale, N.J.: AFIPS Press, 1971. 404 p.

This workshop was devoted to problems and prospects for more effective systems design of the user interface. A challenge paper and several papers prepared in response are included. The discussions are very interesting. Operating experiences from a variety of on-line interactive search systems are discussed, with special regard to the user-computer interface from the user's perspective.

Wiederkehr, R. R. V. "Part I: The Literature Perspective," in Evaluation of Document Retrieval Systems: Literature Perspective, Measurement, Technical Reports. Bethesda, Md.: Westat Research, Inc., December 1968, 1-15. (PB-182 710).

This is a good survey of the literature of the 1960's on the evaluation of information retrieval systems. This report was one of two volumes comprising the first draft of the 1971 King monograph described earlier.

Williams, J. H., Jr. "Functions of a Man-Machine Interactive Information Retrieval System," Journal of the American Society for Information Science 22:5 (September-October 1971) 311-317.

Describes the BROWSER system of IBM Federal Systems Division. The BROWSER system uses a free-form query and produces weighted output. Specific terms (with few postings) are associated with higher values; terms with many postings have lower values.

IV. BACKGROUND INFORMATION

A. THE ERIC DATA BASE

In the mid-1960's the U.S. Office of Education established the Educational Resources Information Center (ERIC) to provide access to literature in the field of education. Through the long-term support of the Office of Education, and currently the National Institute of Education, ERIC has grown to become one of the leading social science information resources in existence today.

To acquire and select material for inclusion in the ERIC data base a network of clearinghouses was established (presently 16 clearinghouses), each with special expertise in a particular area of education. The clearinghouses compile bibliographic information about each publication selected, index each publication using a controlled vocabulary of Descriptors, assign Identifiers, and in some cases write an abstract or brief annotation for the publication. The records thus prepared by each clearinghouse are then sent to a central processing center for further processing.

The two basic printed products of ERIC are the *Research in Education* (RIE) journal and the *Current Index to Journals in Education* (CIJE). Both are published monthly. Concurrently, machine readable versions of the RIE and CIJE files are produced on magnetic tape. These tapes are available at nominal cost on a monthly subscription basis to organizations that wish to search the ERIC files by computer. There are presently about 100 ERICTAPE subscribers. Two other files of educational material are also available on magnetic tape. These files, dealing with topics of vocational and technical education and produced and distributed by the ERIC Clearinghouse on Vocational and Technical Education, are *Abstracts of Instructional Materials* (AIM) and *Abstracts of Research Materials* (ARM).

In addition to the RIE and CIJE publications, and the corresponding files on magnetic tape, a variety of printed listings and indexes are published to aid the searcher. Also, the ERIC Document Reproduction Service offers both microfiche and printed copies of all non-copyrighted reports announced in RIE.

A more complete description of the ERIC system including its scope of coverage, products, services, and operational components may be obtained from the reports of the U.S. National Institute of Education.[1-2]

B. SEARCHING THE ERIC DATA BASE BY COMPUTER

With the creation and monthly updating of machine readable files and their ready availability, much searching of the ERIC document collection is now being done using computers, both in batch mode and on-line. (Batch mode is the procedure of submitting one or more independent search requests to be processed by the computer with no interaction between machine and searcher. Typically, in batch mode several days elapse between submission of the search requests and receipt of the computer output. If modification of a batch request is indicated, a similarly long interval is required for receipt of the new results. On-line operation implies an interaction between computer and searcher during the search process that allows immediate feedback of results and immediate modification of the request when desired.) On-line searching of the ERIC data base is made possible through the facilities of central processing centers which provide the required hardware and software for the searcher as well as maintaining the ERIC bibliographic files in direct access oriented machine-readable form.

A survey of the use of ERIC tapes for computer searching is given in a recent report by Embry.[3] A more detailed review of several installations that search ERIC tapes in a batch mode is given in a recent report by Humphrey.[4]

C. LOCKHEED DIALOG SYSTEM

The Lockheed Palo Alto Research Laboratory has, for a number of years, operated an on-line search facility providing access to a variety of bibliographic data bases. Since 1969 the ERIC data base has been included in the files maintained by Lockheed.

As shown in Figure 2, a subscriber to the Lockheed on-line retrieval service has local terminal equipment (remote from the computer), the required telephone communication link, and on-line access to the Lockheed Palo Alto computer and ERIC data base during a service period of approximately 12 hours each working day.

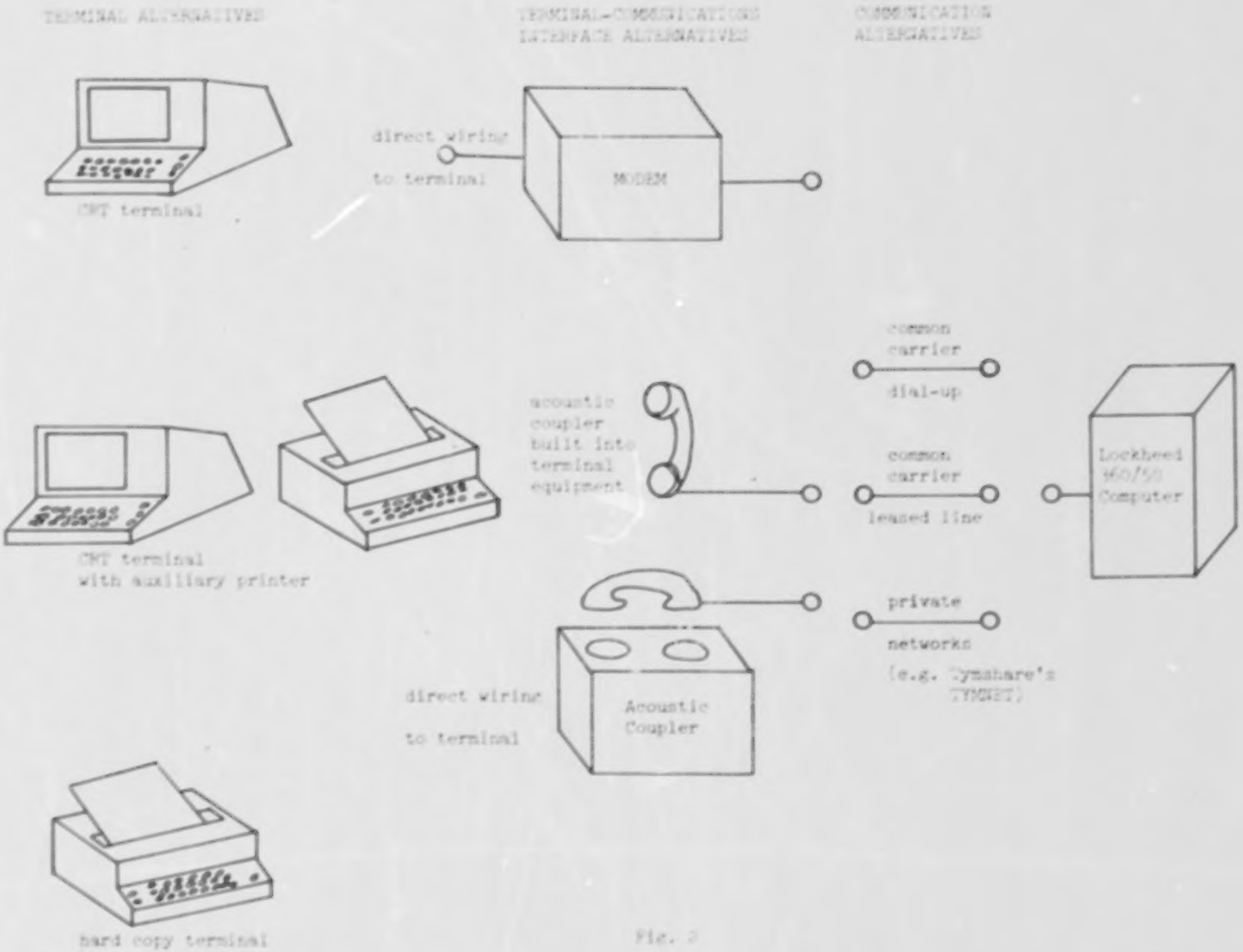
The subscriber's terminal equipment is typically: 1) a cathode ray tube (CRT) video display with keyboard; or 2) a hard copy printing terminal (mechanical, thermal, etc.) with keyboard; or 3) a CRT terminal with an auxiliary printer to print out at the user's option, selected portions of the transmissions.

The terminal equipment and communications channels can both be obtained to handle a wide range of transmission rates. Most of the generally available terminal equipment (mechanical or video) operates at speeds of 10, 15, or 30 characters per second; many models are available that will operate at 120 characters per second, or even 480 characters per second.

Even though some units of the terminal equipment might be able to handle a high data rate, the actual data rate will be limited by the capacity of the communication channel (i.e., the phone line) and the interface to the computer equipment. Most on-line terminals now operate in a dial-up mode in which the subscriber uses an ordinary telephone handset to dial up the number of the computer and then puts the handset into an an acoustic coupler to connect the signal to the terminal equipment. The dial-up arrangement results in the use of whatever telephone line and circuit is selected by the telephone switching equipment; that is, a normal voice grade line, sometimes good--but sometimes not too good. The data rate achievable over such ordinary dial-up lines is theoretically about 1900 characters per second; however, in practice most dial-up terminals operate at a lower rate (seldom more than 120 characters per second) because of considerations of the quality of the data transmission (i.e., problems with available MODEMS, transmission noise and error rates for higher speeds).

Most dial-up users not in the immediate vicinity of Palo Alto currently use Tymshare's TYMNET network to access the DIALOG system. This network provides local phone numbers in over 50 major cities at an hourly connect cost of $10.00. This compares favorably with direct distance dialing which may cost as much as $30.00 per connect hour. TYMNET currently supports terminals in the 10 to 30 characters per second speed range.

Instead of using a dial-up line with its fluctuating quality control problems, some users prefer to lease a line for exclusive use as the connecting link between the terminal and the computer. A leased line, because



TERMINAL ALTERNATIVES
TERMINAL-COMMUNICATIONS INTERFACE ALTERNATIVES
COMMUNICATION ALTERNATIVES
CRT terminal
direct wiring
to terminal
MODEM
CRT terminal
with auxiliary printer
acoustic
coupler
built into
terminal
equipment
common
carrier
dial-up
common
carrier
leased line
Lockheed
360/50
Computer
private
networks
(e.g. Tymshare's
TYMNET)
direct wiring
to terminal
Acoustic
Coupler
hard copy terminal

Fig. 3

Terminal Equipment Input Channel Alternatives Available for the Lockheed DIALOG Installations Studied it is both clearly identified and dedicated to this single application, can be inspected, modified or specially conditioned, and maintained by the telephone company to provide potentially the best performance possible for a telephone line. Such leased lines often operate reliably at speeds up to 480 characters per second. Because the leased line access is provided at a fixed monthly cost with unlimited access (up to 12 hours per day currently), this method has an economic as well as operational advantage for the high level user. Furthermore, because the leased line is always connected to the computer, there is never any difficulty with a busy signal because all of the available lines are being used (i.e., you are never denied a port).

The data that is transmitted over the telephone lines must go through a transformation process between the terminal and the phone lines. Digital signals generated at the terminal must be transformed into equivalent audio analog signals for transmission over the phone lines, and vice versa for transmission to the terminal. This requires some type of MOdulator-DEModulator (MODEM) equipment. Such equipment can be units separate from the terminal equipment (e.g., Bell Telephone Co. Datasets) that take an electrical analog signal from the phone line and directly convert it into an equivalent electrical signal in digital form, or vice versa. However, equipment is also available (acoustic couplers) that will take the telephone audio signal as heard through the handset, and acoustically transform that signal back to an equivalent electrical signal, and vice versa. Using this relatively inexpensive equipment eliminates the need to rent the generally more expensive MODEM equipment. The acoustic couplers are normally purchased as separate units of equipment, but are sometimes built directly into the terminal equipment. The acoustic couplers work quite well at speeds of 30 characters per second, and there are even some units that can operate at speeds up to 120 characters per second.

The computer equipment interface to the telephone lines may also have some restrictions on data transmission rates. The Lockheed computer that was in use at the time of this study had input ports that accommodated transmission rates of 10, 15, 30, 120, and 480 characters per second.

The point of this discussion is to note that there is a wide variety of terminal facilities available, with greatly different characteristics and data transmission rates. The high speed equipment is more expensive, but can process a search faster than the low speed equipment, and may be more cost effective at some moderate level of search activity. All of the installations experienced in this study had somewhat different terminal equipment, and it was expected that this would be related to their search capacity and productivity. A summary of the terminal equipment used at each of the installations studied is given in Table 3.

The Lockheed hardware facilities in Palo Alto consist basically of an IBM 360/50 computer (to be upgraded to an IBM 360/65 in December 1974) with both disc and data cell auxiliary storage with capacity for storing over 5 billion characters of data bases, plus communication equipment to accommodate a large number of remote terminals.

TABLE 3

TERMINAL EQUIPMENT USED BY THE INSTALLATIONS STUDIED

Installation	CRT Terminal Equipment	Hard Copy Terminal Equipment (all operating @ 30 char/sec)	Communication Equipment and Lines	Data Transmission Rate (characters per second)
2	CC-30	GE Terminet300	Modem, leased line	480
3a*	CC-30	GE Terminet300	Modem, leased line	240
3		GE Terminet300	Acoustic coupler, dial-up line	30
4	CC-30	GE Terminet300	Modem, leased line	480
5	CC-30	GE Terminet300	Modem, leased line	480
7	CC-30	GE Terminet300	Modem, leased line	480
9	CC-30	GE Terminet300	Modem, leased line	480
88		GE Terminet300	Acoustic coupler, dial-up line	30
114		GE Terminet300	Acoustic coupler, dial-up line	30
125		GE Terminet300	Acoustic coupler, direct line	30

*The 3a installation operated until July, 1973 and was then transferred to another organization and identified as installation 3. Installation 3 is the one that is analyzed in this report, but 3a data is included in some tables for additional background information.

The subscriber communicates with the Palo Alto laboratory via the DIALOG interactive command language. Messages or commands are entered by the searcher on the terminal's keyboard, and output from Palo Alto is displayed on the CRT screen and/or printed in hard copy at the searcher's terminal depending upon the type of terminal used and the output option selected by the user. The ERIC/DIALOG interactive language consists of approximately 13 basic commands that allow the user to define sets of documents indexed with specified terms or identifiers, to combine defined sets with complete logical flexibility, to browse the ERIC thesaurus, and to select from a variety of output options. It is not the intention of this report to provide a description of the ERIC/DIALOG language. For this the reader is referred to the *DIALOG Terminal Users Reference Manual*,[5] and several other publications that describe the DIALOG system.[6-7]

D. REFERENCES

1. ERIC: A Profile. Washington, D.C. National Institute of Education. Educational Resources Information Center. Undated.

2. ERIC Processing Manual: Rules and Guidelines for the Acquisition, Selection, and Technical Processing of Documents and Journal Articles by the Various Components of the ERIC Network. National Institute of Education. Educational Resources Information Center. July 1974. 544 p. ED-092 164.

3. Embry, Jonathan D., Wesley T. Brandhorst, and Harvey Marron. Survey of ERIC Data Base Search Services. Washington, D.C. National Institute of Education. Educational Resources Information Center. July 1974. 29 p.

4. Humphrey, Allan. Survey of Selected Installations Actively Searching the ERIC Magnetic Tape Data Base in Batch Mode. Vol. I. Berkeley, Calif.: Univ. of California. Inst. of Library Research. June 1973. 86 p. ILR-74-003.

5. DIALOG Terminal Users Reference Manual. Palo Alto, Calif.: Lockheed Palo Alto Research Laboratory. Lockheed Retrieval Service Information Systems Laboratory. n.d., var. paging.

6. Summit, Roger K. "DIALOG and the User -- an Evaluation of the User Interface with a Major On-Line Retrieval System," in Donald E. Walker, ed., Interactive Bibliographic Search: the User-Computer Interface. Montvale, N.J.: AFIPS Press, 1971, 83-94.

7. Summit, Roger K. ERIC On-Line Retrieval System. Use of the DIALOG On-Line Information Retrieval System with ERIC Research in Education Files. Final Report. Palo Alto, Calif.: Lockheed Aircraft Corp. April 1970. 58 p. ED-040 592.

V. ANALYSIS OF TERMINAL INSTALLATIONS

A. INSTALLATIONS STUDIED

The number of installations that search the ERIC data base on-line using DIALOG varies from month to month and from hour to hour. As described later in this report, an important element of our investigation was a detailed examination of all ERIC/DIALOG search activity carried out by nine terminal installations during 15 selected days during October and November 1973. There were a few other organizations that were conducting ERIC searches using DIALOG during this same time period. However, these were not included in the detailed examination because they were not among those suggested by OE for study.

Data was collected for each of these installations by site visits, telephone discussions, and analysis of Lockheed computer records.

B. LOCAL CHARACTERISTICS

Five of the installations included in this report were visited in late 1973 by one of the staff members of this project. The site visits were made for the purpose of obtaining an appreciation of the ways the various centers operated, and also in order to note any distinguishing features which might influence terminal productivity. Several local factors could conceivably affect productivity such as the number of staff members trained in terminal searching, subject training of the searchers, previous experience of the searchers, the length of time that the center had been in operation, and a cost-recovery pricing policy for the users. These and other questions were considered for each of the sites visited.

A brief description of the setting and characteristics of each of these sites is given below. Some of this data is summarized in Table 4. The only characteristics that were common to all of these installations visited were: 1) all of the searching staff were professionals; 2) all of these centers had CRT terminal equipment.

When considering the information in this table and in this section, it should be kept in mind that the data reflect the situation as of late 1973, and may not necessarily describe those centers at the present time.

1. Terminal 2

This installation employs three full-time searchers. This center has been doing on-line searches since 1970. Currently their clientele consists primarily of employees of the U.S. Department of Health, Education and Welfare and also some people from other federal agencies. They use the SEARCH SAVE feature extensively and have been building their own SEARCH SAVE index since June 1973.

2. Terminal 4

This center has one full-time searcher, and is unique as a searching facility in a number of ways. It is an ERIC clearinghouse and has its own in-house abstractors, and searches are run primarily using their own Thesaurus. Because the abstractors are nearby, they can be consulted by the searcher to help resolve problems regarding the way concepts might actually be indexed. Most queries can be handled with two concepts and only one or two terms per concept. A search generally produces 30-60 citations. A duplicate copy of each search is maintained so that if the same query comes in before the next file update, the duplicate can be sent out without having to run the search again.

3. Terminal 5

Terminal 5 has 11 part-time searchers, eight of whom have had considerable experience as school teachers. The staff members are generally quite familiar with ERIC. Their policy is complete and thorough

BEST COPY AVAILABLE

TABLE 4

CHARACTERISTICS OF INSTALLATIONS STUDIED

	Terminal 2	Terminal 4	Terminal 5	Terminal 7	Terminal 9
CHARACTERISTICS OF SEARCH STAFF					
Number of people on search staff	3	1	11	14	2
Staff full or part time on DIALOG	full	full	part time	full & part time	full
Staff had previous experience with ERIC	yes		yes		no
Staff members have advanced degrees in:					
a. education		no	8 staff members average 6 years teaching experience	Most have adv. degrees in ed. or other subjects	no
b. information or library science		M.L.S.	no	2 M.L.S.	M.L.S.
c. other fields		no	no		no
MODE OF OPERATION					
Question asked by client in person, by mail, phone	phone mostly	mail only	mail (mostly)	mail, phone	person & phone
Search strategy worked out before going to terminal--complete or sketch	yes	yes-sketch	yes--complete	not much--sometimes a sketch	yes--sketch
Use primarily descriptors, or both descriptors & identifiers	both	little use of identifiers	both	use primarily descriptors	
Questions automatically go to DIALOG unscreened (i.e., none searched manually or referred elsewhere)	screened	screened for duplicate	about 50% screened out	some screened for manual search	not screened
Check for earlier SAVED searches--always, usually, occasionally	yes	check for earlier duplicate search	not usually	yes	sometimes
Use of EXPAND command	minimize use	minimize use	generally used	use for SELECTing	yes
Availability of analyst tools	ready access	ready access	ready access	at each desk	ready access
Limit number of citations printed--how many	yes if large (varies)	yes 50-100	Yes--about 150 (100 for grad students)	yes 6-10	No-call client if too large
Output from Lockheed reviewed before sending out	yes	not much	yes	yes, using REMKARD, choose 6-10	yes
Hardcopy or fiche sent automatically	varies	no	no	yes, 6-10	no
Hardcopy or fiche sent upon request when possible	varies	yes	yes	sent automatically	yes
Use of intermediaries in the field	no	no	sometimes	usually	no
HISTORY, FUNDING, ETC.					
Year center was started			1967		1973
Year center started using DIALOG	1970	1971			8/1973
Receives funds from NIE	yes	no	partial funding	no	yes
Charges clients for services	no	yes	contracts with area school districts	contract with school systems	free for New Haven, $15-25
Nature of clientele	mostly H.E.W. employees	spec. ed. tchers.; civic groups; grad students	Classroom teachers, school administrators State Ed. consultants	primarily educators in the field	Classroom teachers; administrators; State Ed. Consultants
Other equipment	no	no	Cameras for fiche production & reproduction	REMKARD; TWX; mapcard; hotline	no
OTHER PERTINENT CHARACTERISTICS	Scheduled terminal time same regardless of no. of searches	Search primarily CEC thesaurus; keeps dup. for 3 mo.; simple questions	Searchers often work together; Manual srch to augment DIALOG; Searchers have specific window (30-45 min)	Telecom. syst. for requests; aim for hi precision, low recall	2 types of search: Standard bibliog. & in-depth (includes manual w/DIALOG)

service, including use of manual searching to augment DIALOG when appropriate, and sending hard copy or microfiche reproductions of cited material upon request. DIALOG output is thoroughly reviewed for relevance; even the original documents, as well as ERIC abstracts, are considered in making relevance judgments. The order of priorities in reducing the size of bibliographies which are too large (100-150 being the usual limit) is: 1) limit search to major terms only, 2) limit to CIJE references only, 3) limit by date of document acquisition by ERIC. This installation sometimes uses intermediaries in the field to relay queries.

4. Terminal 7

This center employs 14 searchers (some full-, some part-time), most of whom have an advanced degree in education or another subject area. This center automatically sends its clients hard copy or microfiche reproductions of six to ten items cited in the DIALOG bibliography. Searchers improve their on-line searching speed by making frequent use of EXPAND-SELECT combinations using the chaining method of entering commands (described in a later section), and by doing most of their document screening activities off-line (by using the REMKARD microfiche storage device). Since they are searching for a few "most relevant" documents they search primarily descriptors, use the LIMIT/MAJOR feature extensively, and generally aim for high precision and low recall. This factor might be expected to result in less on-line search time spent per question. When appropriate, DIALOG searches are augmented by a manual search in their library of current publications. This center makes extensive use of intermediaries in the field to report queries in natural language.

5. Terminal 9

Terminal 9 is the newest of the centers visited by the ILR staff, having begun on-line searching just three months prior to the site visit. This factor might be expected to affect the center's productivity for a time. This installation employs two full-time searchers. All searches are done using DIALOG, but a DIALOG search may be manually augmented if an in-depth search is requested. This center's services are free to clients in its local area but those outside the area are charged $15 for a regular search (DIALOG bibliography only) and $25 for an in-depth search. When DIALOG sets are too large, rather than limiting searches to a definite maximum output, clients are telephoned and search questions re-negotiated.

VI. ANALYSIS OF ERIC/DIALOG USE

A. INTRODUCTION

Several steps were taken to explore the basic question of this study, "Why are there wide-spread variations in questions processed per hour across installations?" The project pursued the following major sequence of supporting studies:

1. An investigation of ERIC/DIALOG system response time;
2. A detailed examination of searching patterns of nine installations as provided by a special computer log ("trace histories") of individual DIALOG commands executed by these nine terminals during search operations;
3. A classification of questions processed by the nine installations, according to complexity;
4. A review of the operating policies and procedures of the major users of ERIC/DIALOG during the time period investigated;
5. Analysis of the data obtained.

To aid our investigation, an ERIC/DIALOG terminal was installed at ILR from August, 1973 to March, 1974. The project staff used the terminal extensively during this period and on the basis of this experience, and discussions with searchers from other installations, formulated some general search guidelines which are reported in a later section of this report.

B. STUDY OF ERIC/DIALOG SYSTEM RESPONSE TIME

1. Introduction

The first step carried out in this investigation was a detailed study of the response time of the on-line computer system. This was done to explore the hypothesis that the average response time and resulting search rate experienced by a given terminal might be affected by peak loading of the computer system. That is, terminal use at the busiest days or hours of computer use might experience a slower search rate. Consequently, any terminal installation that, because of local scheduling or East-West Coast hours of operation, tended to come on at the peak hours might experience a systematic lowering of its average search rate. For example, if the system response were significantly faster between the hours of 5:00-7:00 AM then the users on the East Coast might be expected to have shorter elapsed times on the average than those on the West Coast who come on the system at 8:00 AM. (All times used in this report are PST, local California times.) In this experiment, therefore, an attempt was made to measure the system (equipment, communications, programs) response time only; other time which might normally be spent at the terminal--such as time spent thinking about what command to enter next, or reading items on the screen--was reduced to a minimum.

This data collection task aimed to find out whether the system response time (day of week or time of day) could possibly be a factor in the differing average search rates experienced for each of the terminal installations.

We were also interested in determining the difference in search speeds that might be due strictly to the typing speeds and other mechanical skills of the terminal operator.

2. Methodology

Following a review of the command histories of the 9 ERIC/DIALOG terminal users being studied, one fairly representative search was selected. This particular search, shown in Figure 3, was of moderate length and complexity (nearly 50 commands) and makes use of nearly all of the DIALOG commands (i.e., EXPAND, SELECT, RECALL, PAGE, DISPLAY SET HISTORY, DISPLAY ITEM, COMBINE using AND, OR and NOT logic and PRINT). The recorded elapsed time for this original search was 34.22 minutes. (This figure, of course, does include time spent at the terminal thinking, reading displays, etc.)

The chosen search was run repeatedly and continuously during the entire duration of the Lockheed system availability for seven days. This was done from 5:00 AM till 1:30 PM by four searchers on the following days: Thursday Oct. 18, Thursday Oct. 25, Friday Oct. 26, Monday Oct. 29, Tuesday Oct. 30, Wednesday Oct. 31, and Thursday Nov. 1. In this way, data was collected for each day of the week and on one day (Thursday) for each of three consecutive weeks.

BEST COPY AVAILABLE

```
1                                  (SELECT File 1--ERIC)
.RECALL 5V                         (RECALL a saved search)
"ADULT EDUCATION                   (EXPAND ADULT EDUCATION)
"E6                                (EXPAND line reference E6)
#R1                                (SELECT line reference R1)
#R4                                (SELECT line reference R4)
#R5                                (SELECT line reference R5)
#R7                                (SELECT line reference R7)
0                                  (PAGE forward in Expand display)
#ADULT EDUCATION PROGRAMS          (SELECT ADULT EDUCATION PROGRAMS)
@                                  (DISPLAY SET HISTORY)
$1-5/+                             (COMBINE sets 1 to 5 with OR)
"FEDERAL AID                       (EXPAND FEDERAL AID)
"E6                                (EXPAND line reference E6)
#R1                                (SELECT line reference R1)
#R3                                (SELECT line reference R3)
#R6                                (SELECT line reference R6)
#R7                                (SELECT line reference R7)
#R12                               (SELECT line reference R12)
"GOVERNMENT ROLE                   (EXPAND GOVERNMENT ROLE)
                                   (NULL command)
"GOVERNMENT ROLE                   (EXPAND GOVERNMENT ROLE)
#E6                                (SELECT line reference E6)
                                   (NULL command)
@                                  (DISPLAY SET HISTORY)
$7-12/+                            (COMBINE sets 7 to 12 with OR)
$6*13                              (COMBINE sets 6 and 13 with AND)
"EVALUATION                        (EXPAND EVALUATION)
"E6                                (EXPAND line reference E6)
#R1                                (SELECT line reference R1)
#R16                               (SELECT line reference R16)
0                                  (PAGE forward in Expand display)
0                                  (PAGE forward in Expand display)
#R36                               (SELECT line reference R36)
#EVALUATION CRITERIA               (SELECT EVALUATION CRITERIA)
#FOLLOW UP STUDIES                 (SELECT FOLLOW UP STUDIES)
#REPORTS                           (SELECT REPORTS)
#ANNUAL REPORTS                    (SELECT ANNUAL REPORTS)
$15-21/+                           (COMBINE sets 15 to 21 with OR)
$14*22                             (COMBINE sets 14 and 22 with AND)
%23                                (DISPLAY set 23)
0                                  (PAGE forward in Display)
0                                  (PAGE forward in Display)
&23/5                              (PRINT set 23 in format 5)
&                                  (Continue PRINT of set 23)
#SECONDARY EDUCATION               (SELECT SECONDARY EDUCATION)
#SECONDARY GRADES                  (SELECT SECONDARY GRADES)
#POSTSECONDARY EDUCATION           (SELECT POSTSECONDARY EDUCATION)
$24+25+27                          (COMBINE sets 24, 25, and 27 with OR)
$14*28                             (COMBINE sets 14 and 28 with AND)
$29-23                             (DELETE all items in set 23 from set 29)
&30/5                              (PRINT set 30 in format 5)
=                                  (END)
```

Fig. 3. Query Used for the DIALOG System Timing Exercise

An attempt was made to run the search line-for-line exactly as it was originally run except for two minor changes. PRINT format 6 was used instead of format 2 in an effort to save paper. Also, the title, searcher, requestor and address were changed to suit the purposes of the experiment. (The title space was used to record the start time of the search.) No time was spent reading displays or doing anything other than simply entering the appropriate command. Each command was entered as soon as possible after the blue keyboard light and the "Enter" signal appeared. (There were some exceptions to this general rule as discussed in the later section on possible sources of error.) The searches were done as fast as possible without straining the searchers. Usually, one searcher completed two or three searches before another searcher took over, consequently fatigue and boredom probably did not significantly affect the results of the experiment.

Each searcher recorded the real clock time at start and finish of each search as well as the elapsed time indicated by the system at the end of each search. (The real elapsed time corresponded with the system report except for searches during which the system was down.) Notes were also made regarding any peculiar behavior on the part of the system, when the system went down (if it did), any significant interruptions of the searcher's work by telephone calls or other distractions, and keyboard flubs (if any) made by the searcher.

3. Possible Sources of Methodological Error

Before discussing the results of this effort, a number of possible sources of error should be pointed out.

On Thursday Oct. 25 the RECALL command was unavailable from 8:45 AM through the rest of the day. It was agreed that the searches would continue anyhow since the recalled search was not actually used in this search but merely called up and rejected. It was assumed that the amount of time it would take the system to respond with a display of the saved search would not be very different from the time it would take the system to respond with the "Invalid command" message. Therefore, the searches were continued and the RECALL and PAGE commands were entered at the appropriate places in the sequence.

Another minor deviation from the ideal methodology was that one of the searchers did not wait for the "Enter" signal to appear on the screen but started keyboarding the command as soon as the blue light indicated that the keyboard was available for use. (This blue light usually comes on for a few seconds then goes off and comes back on again simultaneously with the "Enter" signal.) The other three searchers waited till the light appeared the second time before starting to enter the commands. Probably the only place where this difference in method would affect search time is when entering commands which include whole words which must be typed in. It was thought that the results of this discrepancy in method of entering commands would be negligible and would merely tend to offset individual differences in typing speeds.

Possibly the most serious source of error lies in inconsistencies in the tolerated degree of divergence from the original search. In general it was agreed among all searchers that if an imperfect command was accidentally entered early in the search such that the following set numbers would not correspond with those of the original search, then that search should be aborted and a new one begun. But if an imperfect command was entered near the end of the search such that any resulting difference in elapsed time would be negligible, then the search should be completed and included in the data. Obviously this is a subjective judgment and may therefore vary both between individual searchers and for each searcher at different times. It has been assumed that the effects of such differences in judgment will be averaged out in the results of the data. Thus, only one completed search was ignored in the results pertaining to system response time. This was the first search done by one of the searchers which included a number of incorrect commands and resulted in an unusually high search time (25.17 min.). This search is included in analyzing the data according to different searchers but not when analyzing the data according to differences in system response at different times of the day since this search does not truly reflect a difference in system response time.

Usually each searcher completed two or three searches before another searcher started. However, on Monday Oct. 29, only one searcher was available until 9:40 AM and therefore completed fifteen consecutive searches. The elapsed times recorded during that time were, however, consistently low so any effects of fatigue or boredom were apparently negligible.

4. Results

Table 1 in Appendix A shows the searcher, starting and ending times, and system-recorded elapsed time for each search completed during each of the seven days. This data was analyzed according to the response times of individual searchers, and system response times at different times of day and different days of the week.

In Table 2 of Appendix A the data have been arranged according to searcher for each day searches were done. The cumulative mean elapsed time for each searcher from the first search done by that person on the first day through the last search done by that person on the last day is also indicated. The mean elapsed time and standard deviation of the sample are given for both the daily sample and the entire series of searches for each searcher.

Figure 4 shows the pattern of search times for all searches done by each of the four searchers. The search times were plotted sequentially from the first search on the first day to the last search on the last day. In these graphs, searches which overlapped with system down time were excluded. Figure 5 shows plots of the cumulative mean elapsed times for each searcher; down-time searches were excluded here also.

In Figure 6 the curves for each day's searches are plotted, including down-time searches (indicated as such). It can readily be seen that most of the peaks represent searches that were interrupted by system down time.

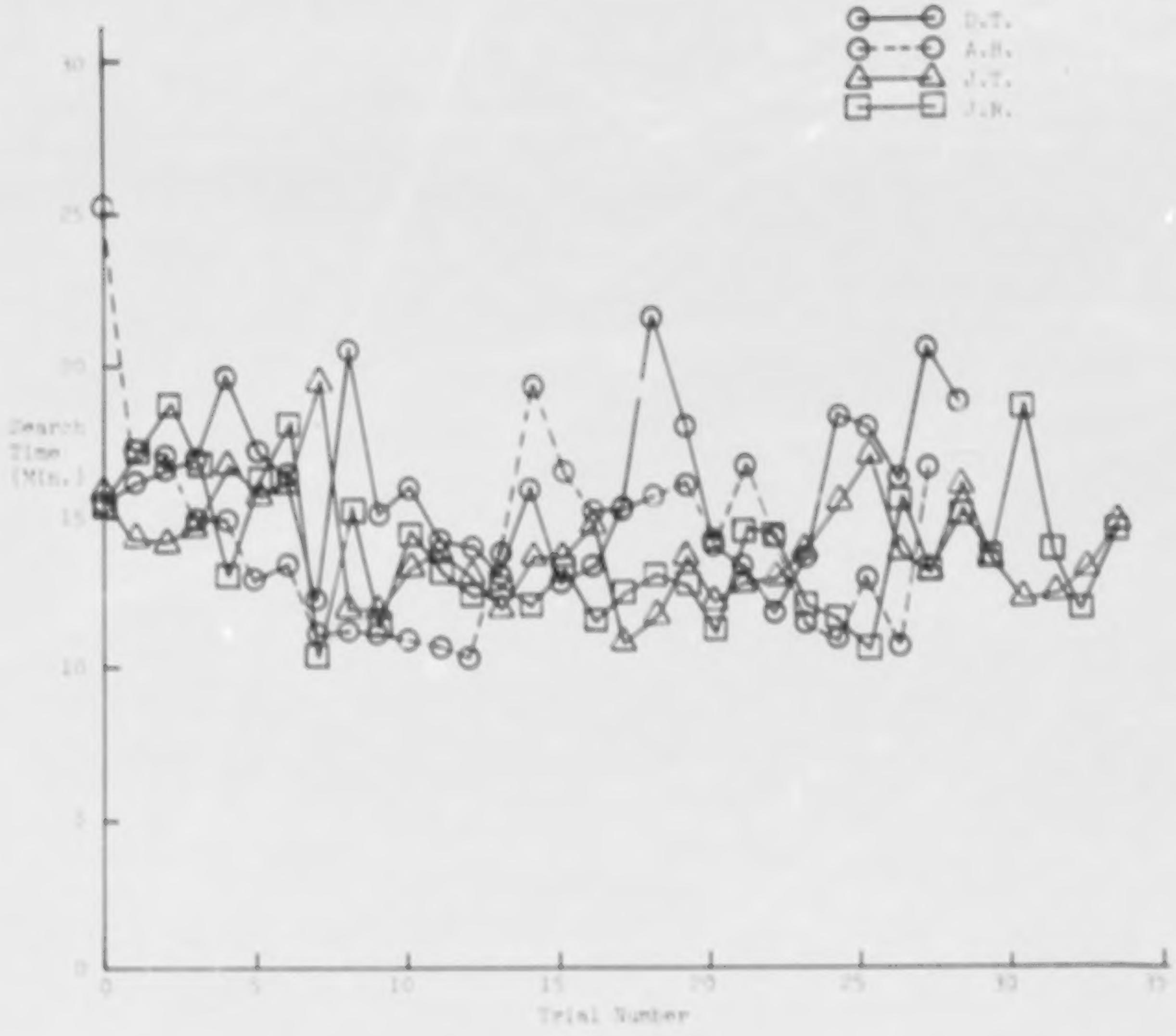


Fig. 4. Search Times for All Searches Done by Each Searcher

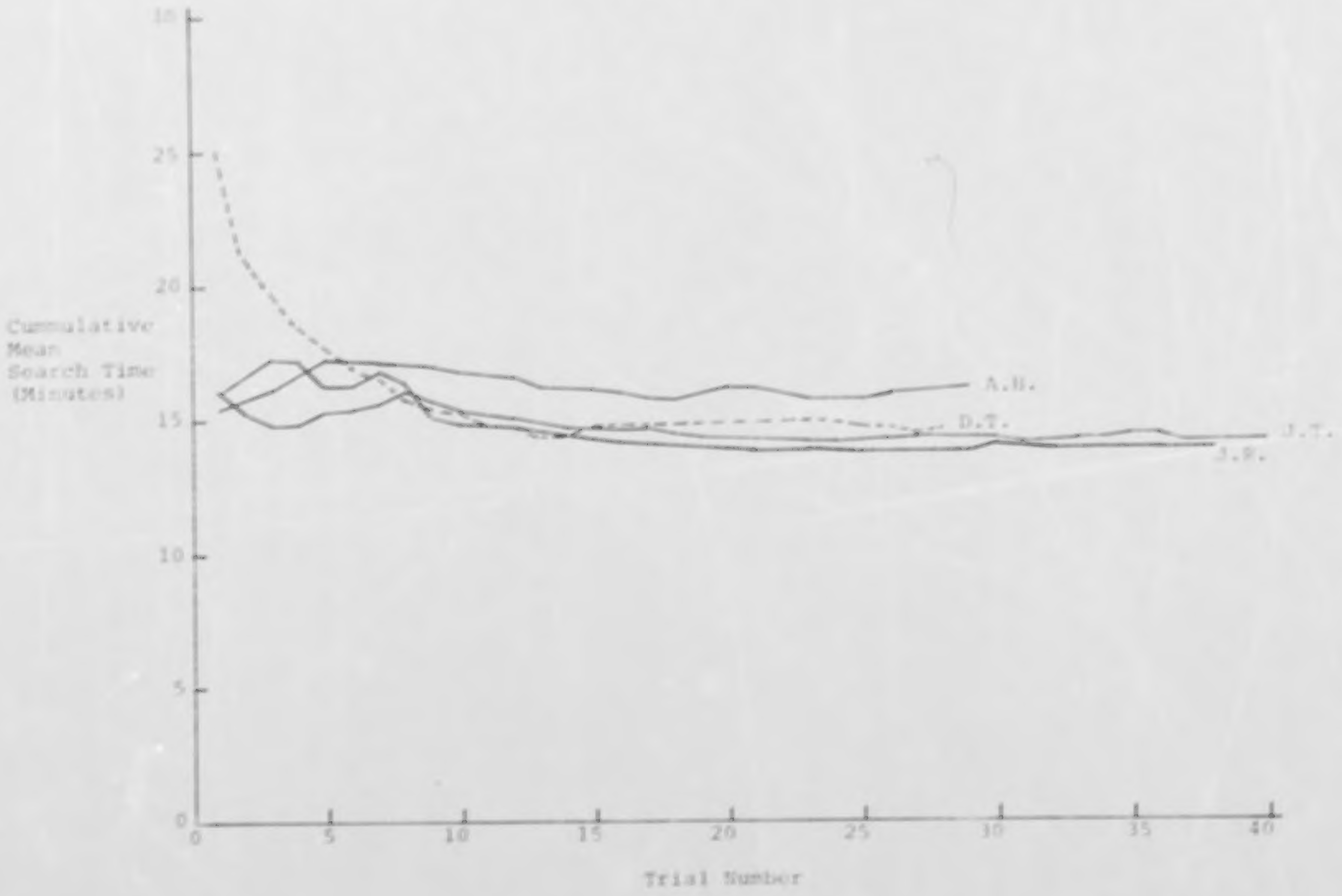
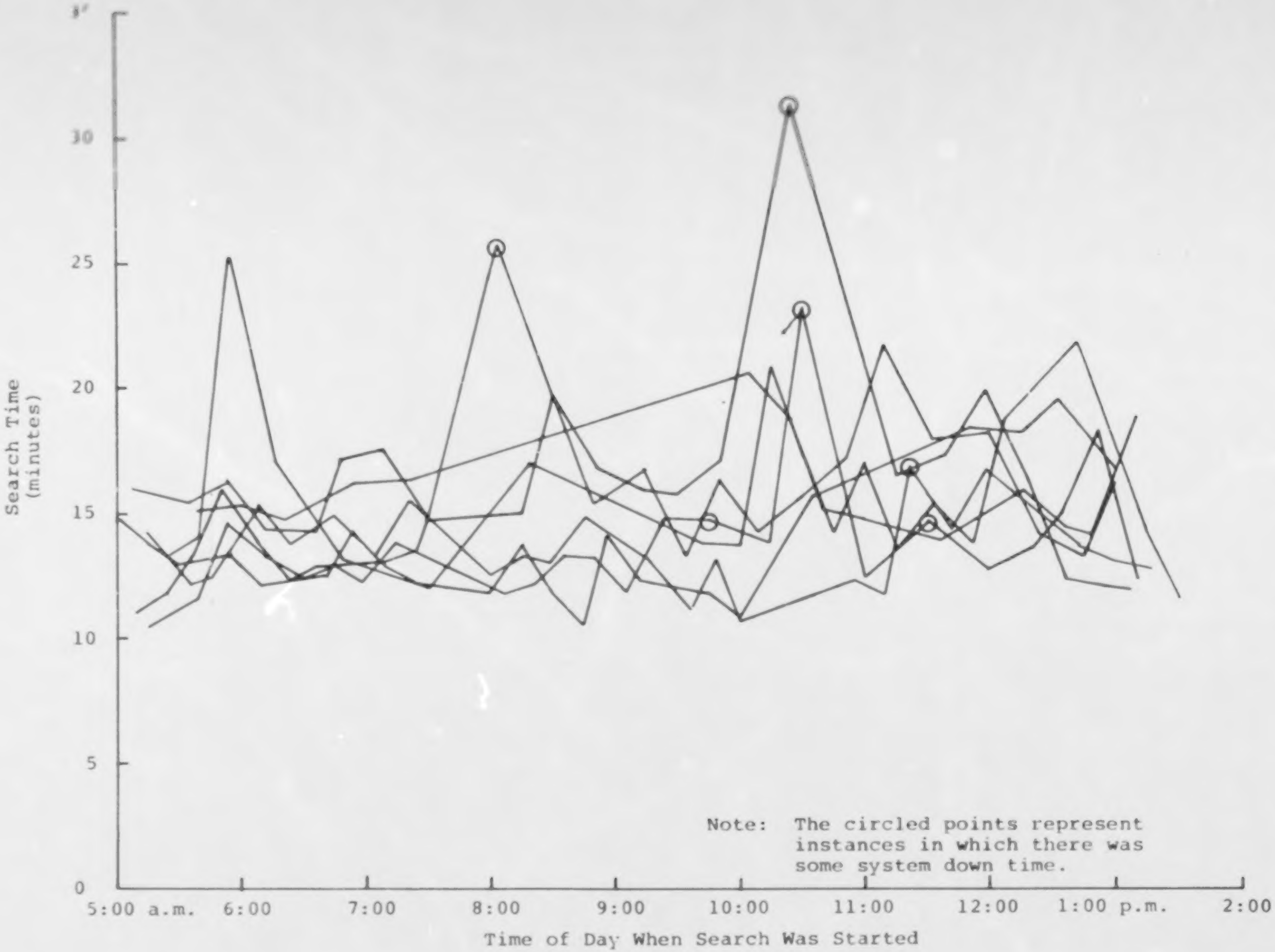


Fig. 5. Cummulative Mean Search Time for Each Searcher



Note: The circled points represent instances in which there was some system down time.

Fig. 6. Sequence of Search Times for Each Day of Timing Searches

The next two graphs (Figure 7 and Figure 8) are envelope graphs spanning the shortest and longest search times for various times of day for the seven-day period. Figure 7 excludes down-time and Figure 8 includes down-time searches.

The data were separated into hourly periods such that all searches that started between 5:00 and 5:59 were grouped together, all those starting between 6:00 and 6:59 were grouped together, and so on. Figures 9 and 10 plot the mean search time for each hourly group, with Figure 9 excluding, and Figure 10 including down-time searches. Figure 11 plots the mean search lengths for each day of the week.

5. Conclusions

a) Influence of Search Operator's Keyboarding Skills

Our searchers possessed a range of keyboarding skills. One searcher had almost no typing skills, while the rest of the searchers had at least an average typing skill. Although there was some individual variation in response time among searchers, this was considered slight enough not to bias the results of the experiment. Individual searchers were relatively consistent in their search times for the same question, seldom varying more than ±2 minutes for a search of about 15 minutes average duration.

There were systematic differences in the mechanical operating speeds of the four searchers, with overall average times of 13.93, 14.03, 14.56, and 16.21 minutes. However, it does not appear that the mechanical skills of the terminal operator are a significant factor in explaining the time variances of the regular system users. A non-typist can do a search about as fast as a skilled typist.

The cumulative mean elapsed times for each searcher fluctuated little after a learning curve of eight or nine searches. Individual searchers had a relatively rapid learning curve for this effort, with usually only the first few searches being significantly slower than the remaining searches.

For this sample of one real query, the mechanical operations and system time accounted for 15 minutes out of the total recorded elapsed time of 34.22 minutes, providing some indication of the relative importance of the system speed versus the operator's cerebral speed.

b) Influence of Hour of the Day

The curves of response time according to the hour of the day were fairly level, with some tendency to peak at 10:00, 11:00 and 12:00. Using the mean search length for hourly periods, excluding down-time searches, these peaks, as shown in Fig. 9, were only two minutes longer than the shortest average hourly search length. Including down-time searches, the difference between the lowest hourly mean search length and the highest was slightly less than three minutes. From this it could be concluded that the variable of system response time at different times of day does not significantly affect the recorded elapsed times of searches. No data was available

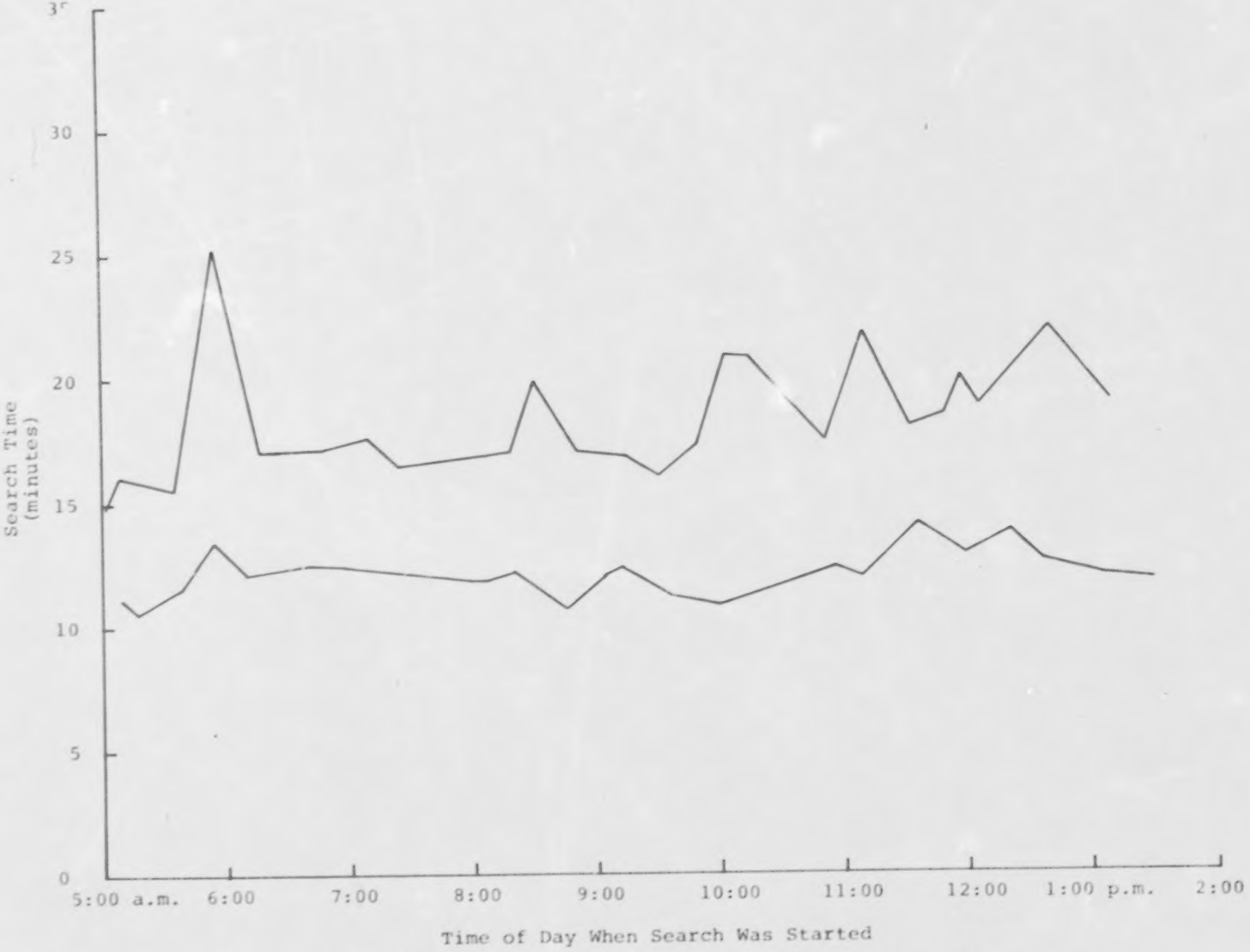
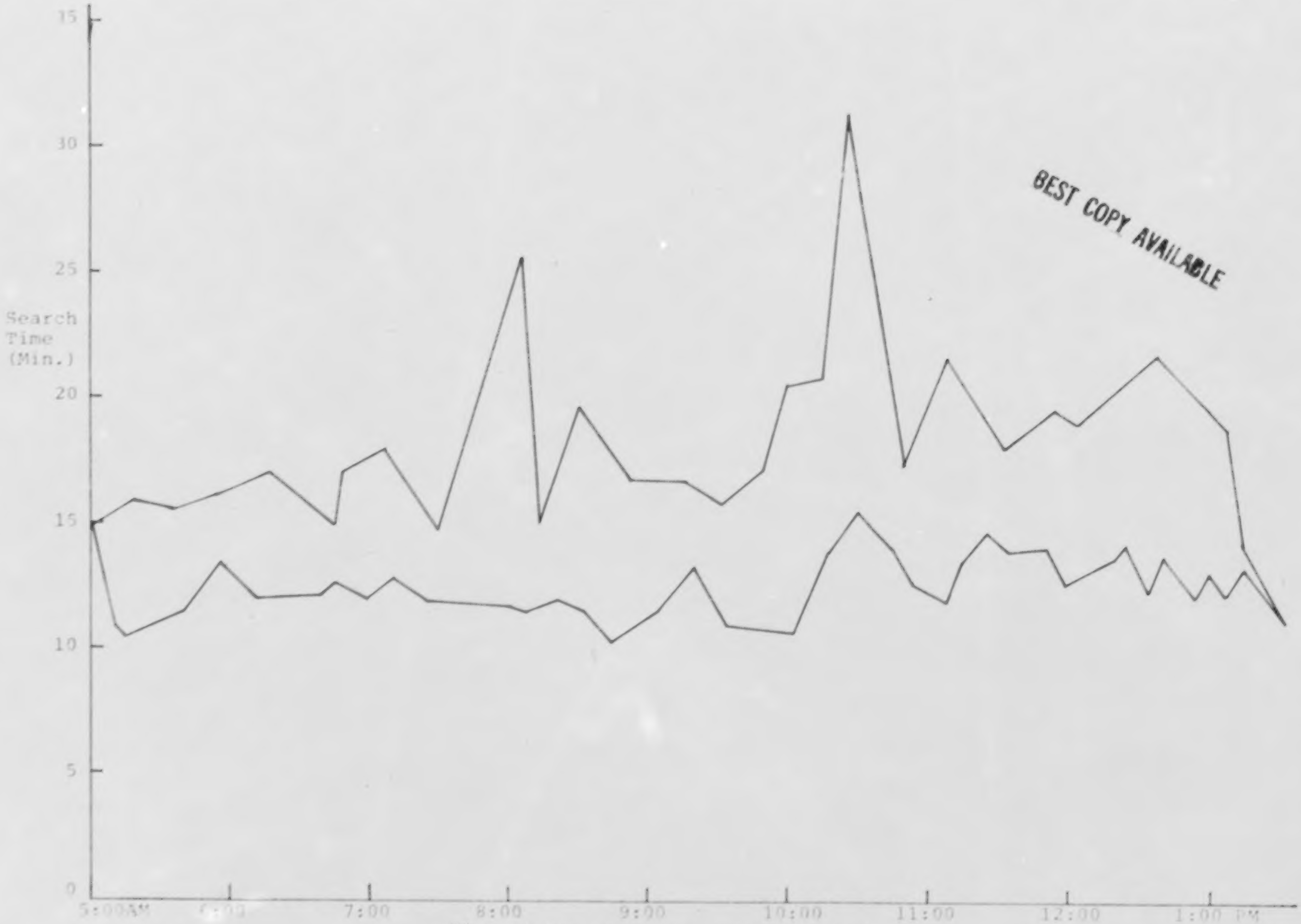


Fig. 7. Range of Search Times Experienced with Timing Searches--Excluding Down-Time Searches (Maximum-Minimum Envelope)



BEST COPY AVAILABLE

Fig. 8. Range of Search Times Experienced with Timing Searches--Including Down-Time Searches (Maximum-Minimim Envelope)

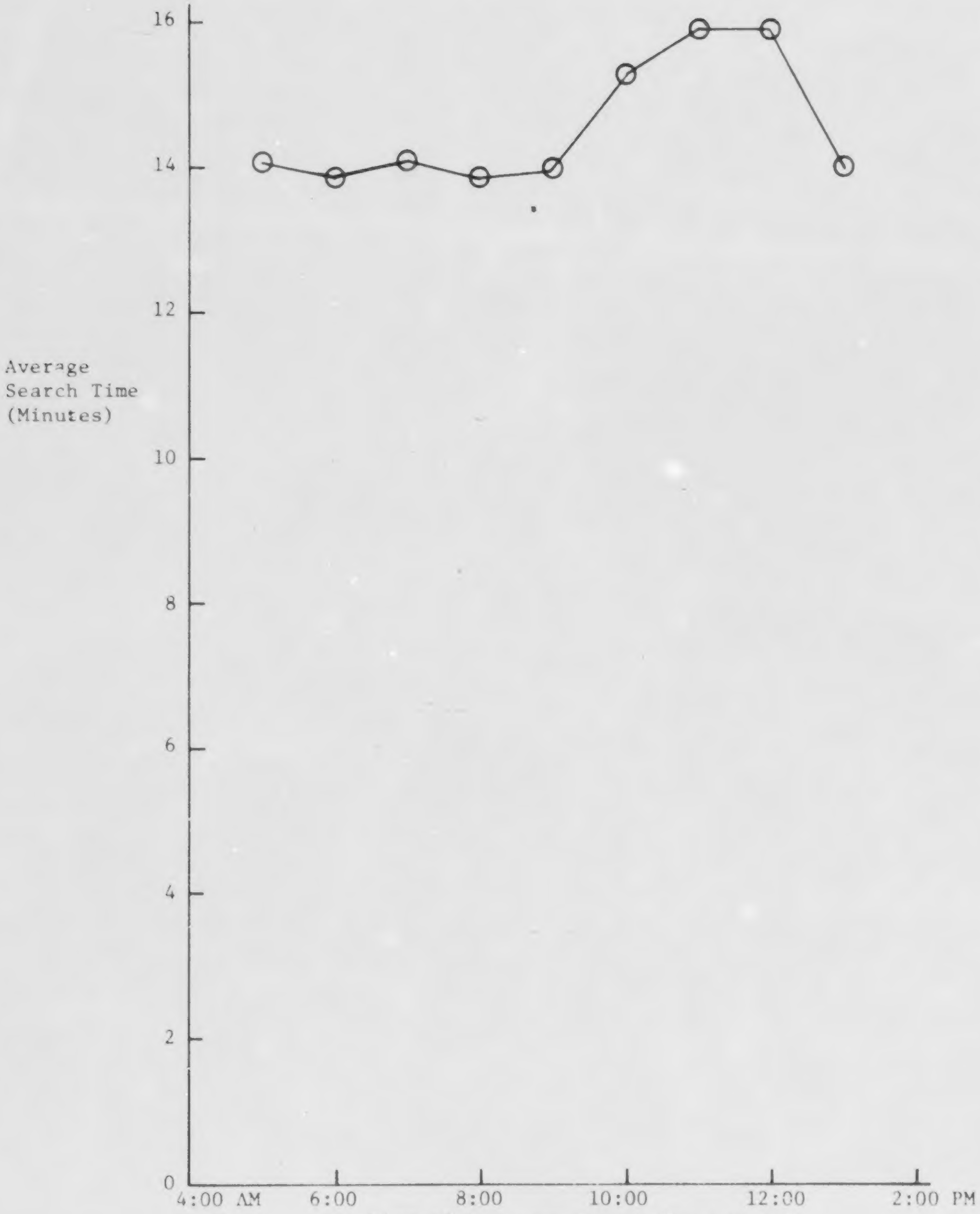


Fig. 9
Average Search Time for All Timing Searches Started at Various Times of the Day
(Excluding Down-Time Searches)

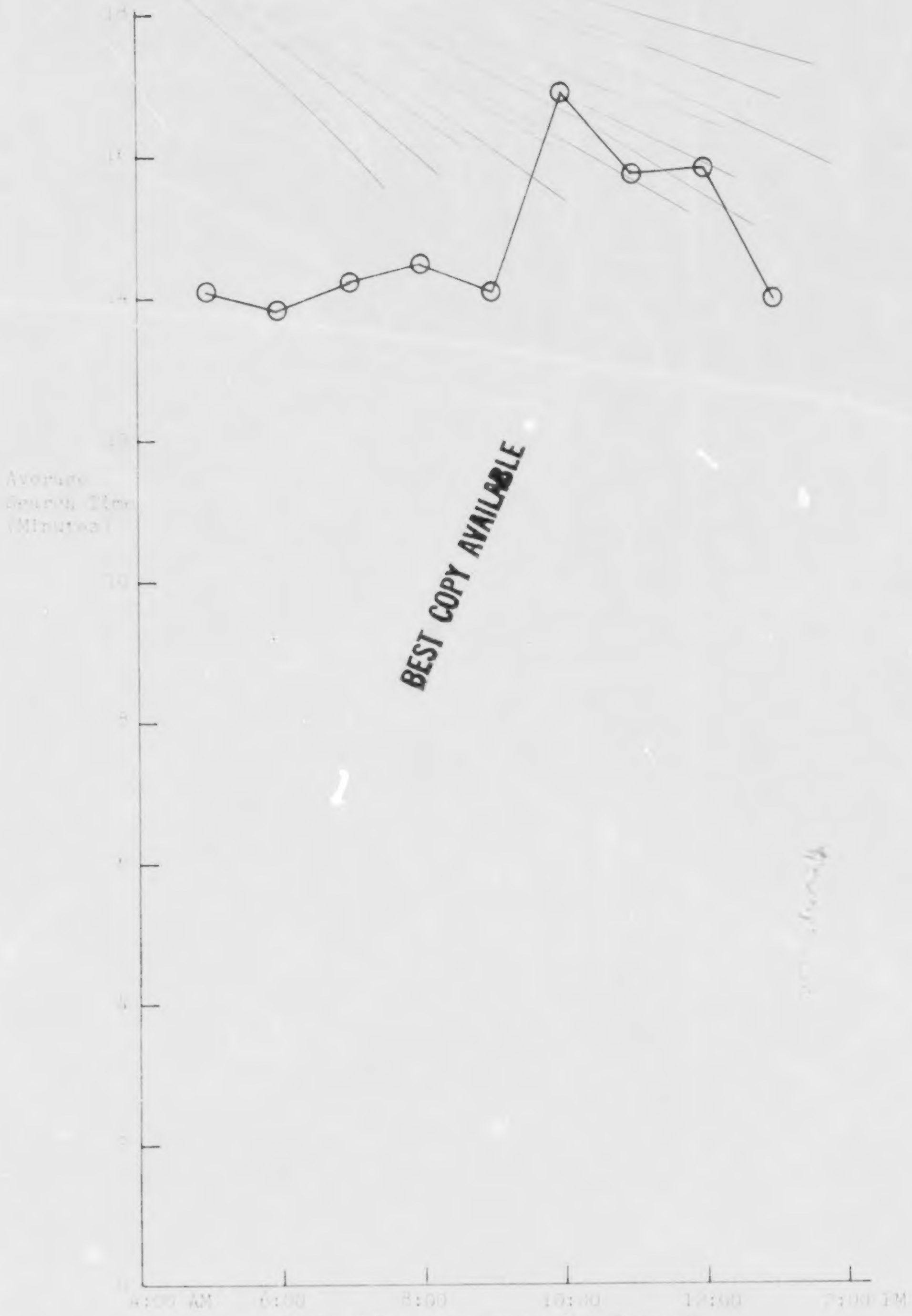


Fig. 10. Average Search Time for All Timing Searches Started at Various Times of the Day (Including Down-Time Searches)

BEST COPY AVAILABLE

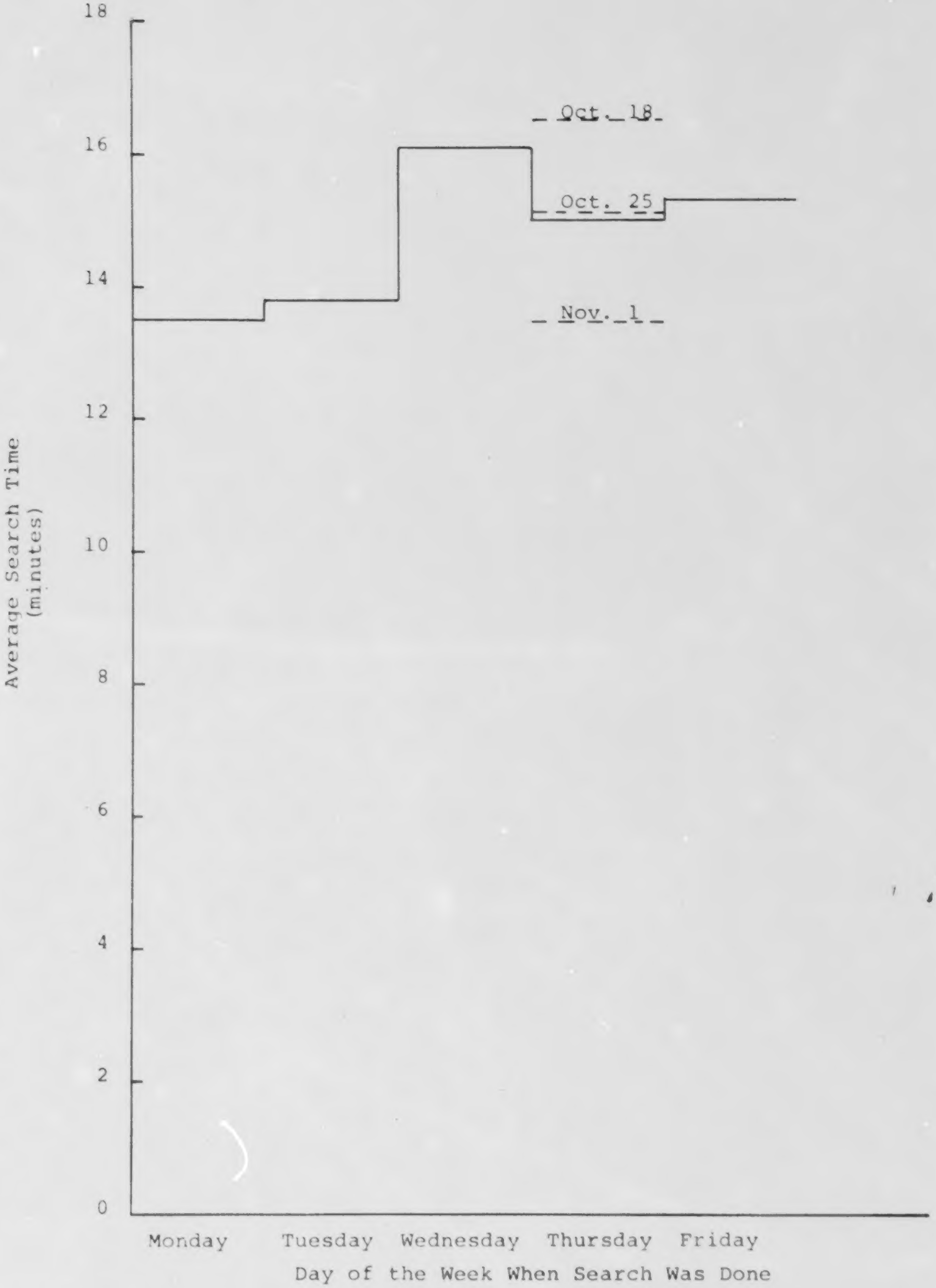


Fig. 11. Average Search Time for All Timing Searches Run on a Given Day of the Week

in the Lockheed records to show what the total volume of computer processing was during the time that our test runs were made. Thus we cannot directly correlate the system response time in our tests to the computer load at that same time.

Average overall times of 14.1, 13.9, 14.1, 14.5, 14.1, 16.8, 15.8, 15.8, and 14.0 minutes were measured for the one-hour time periods beginning at 5:00 AM, 6:00 AM, to 1:00 PM, respectively. However, this difference amounted to about 3 minutes for this 15 minutes search, and thus does not appear to be a significant factor in explaining the time variances of the regular system users. A search will take approximately the same time to complete regardless of what time of the day it is run.

However, a word of caution in accepting this conclusion seems called for: although the elapsed time recorded for a down-time search may not differ greatly from that of searches which did not include down time, the elapsed real time may be quite long. For example, on October 30, a search was begun at 10:37 and concluded at 11:45. The recorded elapsed time was 15.78 minutes; the real elapsed time was over an hour. Since for the purpose of this experiment only system response time was being measured exclusive of any operator "think-time" or goof-off time, our searchers stayed by the terminal in such cases and completed the search as soon as possible after the system came back up. If instead the searchers had left the terminal during down time to attend to some other task--as may often happen in real search situations--and had returned some time after the system came back up (say, 15 minutes later), might this have noticeably affected the recorded elapsed time (e.g., added 15 minutes to it)? The answer to the question may be No, in which case down time may not significantly affect average search length. If the answer is Yes, then the question of whether the system tends to go down more frequently at certain times of the day should be researched. (According to our small sample, the system tends to go down most frequently between 10:00 and 11:00 AM; the next most frequent time is between 7:30 and 8:00.)

Since the overall monthly down time has been measured by Lockheed to range from about 2% to 5%, with the 2% figure being more typical, the overall effect of system down time on productivity should be negligible. Lockheed personnel stop the system clock as soon as an evidence of malfunction is seen, hence system down time is almost completely eliminated from the reported search elapsed times.

One purely subjective caution might also be added pertaining to the effects of system failure. One of the searchers stated that the system "felt" much faster during the first few hours of the day and during the last half hour or so. Although this difference is reflected in a very small amount of elapsed time (at most 2 or 3 minutes), it seems to make a greater psychological difference than the figures suggest. An operator may tend to get impatient during slow or down-time searches, and this may affect the operator's performance at the terminal. For the searchers doing this experiment, there was no "cerebral time" involved. The commands were simply fed in verbatim, so we merely became bored or impatient during slow or down system time. But for "real" operators who have to think about what they're doing, slow or down system time may result in impatience plus disrupted trains of thought. At any rate, there may be a significant difference between the performance of an operator using the terminal when its response is optimal and an operator using it when the response is sluggish. If so this might also affect the average search length of "real" searches.

c) Influence of Day of the Week

There was some variation in mean elapsed times for different days of the week, but this does not appear to be significant. The difference between the shortest mean elapsed time (13.5, Monday) and the longest (16.1, Wednesday) is less than the difference between the shortest and longest mean elapsed times for the three Thursdays (13.5 and 16.6).

C. DETAILED EXAMINATION OF ERIC SEARCHES

1. The Source Data

The DIALOG software has a capability to provide a computer log which records each DIALOG command presented for processing, the identification of the terminal which submitted it, and the date and time (to the nearest second) when it was executed at the central computer. This continuous chronological log of the total operations for the nine terminals in this study was specially prepared to enable us to do a post-audit to trace the precise sequence of commands carried out by each of the nine individual installations. Lockheed provided this selective command log on magnetic tape for 15 selected days of operation during the Fall of 1973. All of the 15 days were ones in which there were no problems with the performance of hardware, software, or communications equipment, i.e., DIALOG had no down time on any of these days. Since timing data was an important part of our analysis it was essential that data be collected from trouble-free days only. The 15 days studied were October 11, 17, 22, 23, and November 1, 5, 6, 9, 14, 15, 19, 21, 23, 26, and 27.

2. Definition of Search and Question

First, it is important to understand the distinction between the term "search" and "question". We have followed the same convention regarding the meaning of these terms as has Lockheed in its periodic activity reports in the DIALOG/CHRONOLOG. The start of a new search is indicated only by submission of a DIALOG "BEGIN" command. A search is considered to be terminated only by encountering in the trace history another BEGIN command or a system-generated message that says I/O SUBTASK TERMINATED (indicating that the terminal has signed off) or a system-generated message (e.g., beginning with the word DIALOG). A question, on the other hand, is considered to be completed if any of the above conditions occur or if a DIALOG "END" command is submitted. It is common practice for a DIALOG user to commence operations at the terminal with a BEGIN command and then proceed to submit several different logical queries or questions, each terminated by an END command, before submitting another BEGIN. Such a sequence would be interpreted to be one SEARCH but several QUESTIONS. The sequence of commands shown in Figure 12 provide a further illustration of this definition.

The data tape initially had the following characteristics:

772	searches
1,129	questions
281	terminal-hours of connect time.

A few command sequences which would be QUESTIONS by the preceding definition but which appeared to represent housekeeping functions or afterthoughts were deleted (manually) from the trace history log. These were generally single commands between END commands (e.g., messages, a single

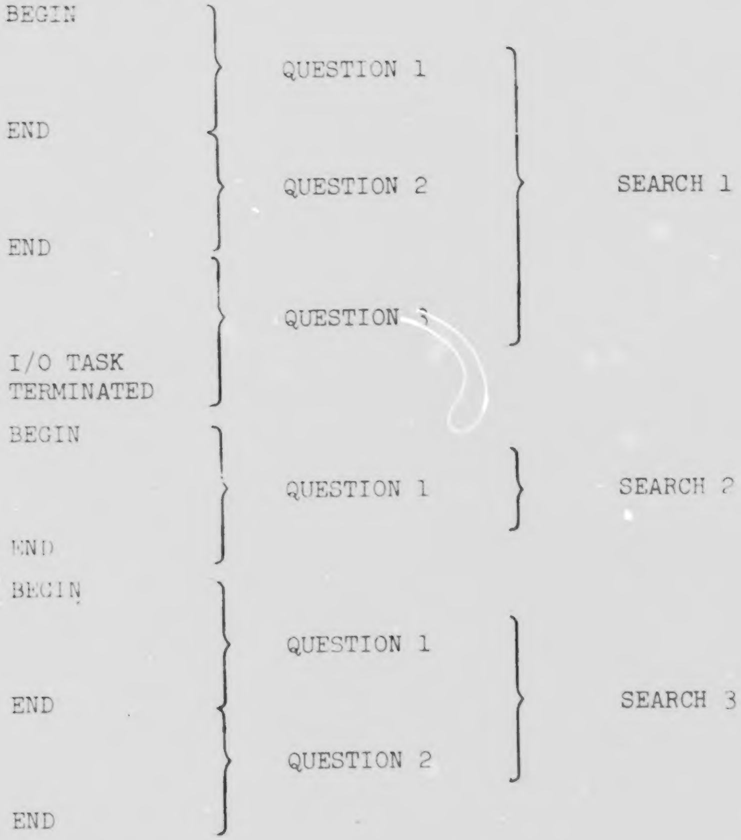

BEGIN
QUESTION 1
END
QUESTION 2
SEARCH 1
END
QUESTION 3
I/O TASK
TERMINATED
BEGIN
QUESTION 1
SEARCH 2
END
BEGIN
QUESTION 1
END
SEARCH 3
QUESTION 2
END

Fig. 12. Illustration of Question and Search Definitions

SELECT with no output function, a single PRINT with no logic function, or a BEGIN-END combination with nothing between).

After this manual editing of the tape the data tape had the following characteristics:

730 searches

1,011 questions

239 terminal-hours of operation.

This data forms the basis for the detailed examination of ERIC searches.

This continuous log on magnetic tape was processed by a sequence of computer programs written by ILR to produce the tables given in this chapter.

The timing data was gathered by subtracting the time at which a given command was executed by the central computer from the time at which the subsequent command was executed. The resulting figure represents think time, keying time, execution time for the first command, transmission time back to the searcher, searcher and system transmission time for the next command to the central computer. The result of scanning the computer log provides reliable timing data on command use, with one exception. It is frequently observed that after termination of a search by an END command, a searcher may get up from the terminal and take a break, coming back to resume searching after 10 or 15 minutes. In such cases, the 10 or 15 minutes would be counted by our program as having been associated with the END command, whereas the time may really have been a between-search pause.

Theoretically, there should be a trivial amount of time associated with the END command. The data analysis programs were run to count the END command time, but the results were subsequently edited to delete the END times and re-distribute the percentages and totals to the amounts now shown in the tables in this section.

One part of the data reduction effort included a count of the number of logical operators (AND, OR, NOT) in each question and search. Because the DIALOG system permits a single COMBINE command to perform several logical operations, it was necessary to locate each of these instances and expand the command to get the proper count of logical operators. For example, the command

```
S1-4/*        (i.e., set 1 AND set 2 AND set 3 AND set 4)
```

is equivalent to

```
S1*2*3*4
```

and has been counted as three AND operators.

Similarly, the command

$7-9/+ (i.e., set 7 OR set 8 OR set 9)

is equivalent to

$7+8+9

and such a command has been counted as having two OR operators.

3. Findings

Table 5 provides summary data (by terminal) for the 15 full days of operation. This includes gross data about the number of questions and searches from each terminal, a tally of the number of times that each logical operator was used with each terminal, and combinations of these parameters. There are some wide terminal-to-terminal variations in these figures, and they are summarized in this table. Notice the approximately 3:1 ratio in the use per question of the AND, 5:1 ratio in the use per question of OR, and 32:1 ratio in the use per question of the NOT operator.

The difference in use of ANDs may reflect different approaches to retrieval for the same question. Most of the installations used between 1 and 3 ANDs per question, with an average of just over 2 ANDs per question. This figure is a little lower than that reported in 1969 by Roger Summit[1] for the RECON system, where the average number of Boolean intersections as used by RECON searchers was 2.30. The data given in 1972 in a related study by Martha Williams[2] of the number of ANDs used in 126 SDI profiles running on the IITRI computer-based current awareness system leads to an estimate of 2.37 ANDs per profile. With the assumption that an average on-line search in our study is equivalent to an average profile in the IITRI study, then this would be 2.37 ANDs per question.

The variation in number of ORs per question (a range of about 2-11 ORs per question) probably reflects some searching installations' practice SELECTing a range of terms from a display with one SELECT command. Terms thus SELECTed are automatically ORed together, and thus fewer ORs are required to be keyboarded to establish the OR relationship. For a further discussion of this point, see tne later chapter on Search Guidelines. The data in the Williams article leads to an estimate of 3.01 ORs per question for their situation.

The great variation in NOTs used per question (.02-.64 NOTs per question) is probably not statistically significant due to the infrequency of occurrence in this small sample (only 154 times in over 1000 questions). The data in the Williams report leads to an estimate of .47 NOTs per profile. A 1971 report by James Carmon[3] of the University of Georgia experience in computer-based current awareness services noted that, "The six profile batches range from 3% to 11% of the profiles which use terms with NOT logic." Use of the NOT logic would seem to be related also to the particular data base being searched.

TABLE 5

QUESTIONS, SEARCHES, AND LOGICAL OPERATIONS PER TERMINAL

	TERMINAL NUMBER												
	(SLOW TERMINAL EQUIPMENT)				(FAST TERMINAL EQUIPMENT)								
NUMBER OF	3	88	114	115	2	4	5	7	9	TOTAL	LOW	HIGH	HI/LOW RATIO
QUESTIONS	31	32	113	54	58	239	112	302	70	1011	31(3)*	302(7)	9.7
SEARCHES	22	31	111	49	43	187	78	144	65	730	22(3)	187(4)	8.5
ANDS	56	52	455	83	139	311	245	551	194	2086			
ORS	193	70	897	140	631	966	702	1047	605	5251			
NOTS	10	2	14	2	37	4	20	31	34	154			
QUESTIONS/SEARCH	1.41	1.03	1.02	1.10	1.35	1.28	1.44	2.10	1.08	1.38	1.02(114)	2.10(7)	2.1
AND/QUESTION	1.81	1.62	4.03	1.54	2.40	1.30	2.18	1.82	2.77	2.06	1.30(4)	4.03(114)	3.1
AND/SEARCH	2.55	1.68	4.10	1.69	3.23	1.66	3.14	3.83	2.98	2.86	1.66(4)	4.10(114)	2.5
OR/QUESTION	6.23	2.19	7.94	2.59	10.88	4.04	6.27	3.47	8.64	5.19	2.19(88)	10.88(2)	5.0
OR/SEARCH	8.77	2.26	8.08	2.86	14.67	5.17	9.00	7.27	9.31	7.19	2.26(88)	14.67(2)	6.5
NOT/QUESTION	.32	.06	.12	.04	.64	.02	.18	.10	.49	.15	.02(4)	.64(2)	32.0
NOT/SEARCH	.45	.06	.13	.04	.86	.02	.26	.22	.52	.21	.02(4)	.86(2)	43.0

* Terminal numbers are shown in parentheses when the value relates to a single terminal.

Table 6 provides a summary distribution of the number and relative percent of DIALOG commands used by each terminal. There are differences here in each terminal's use of these commands. During the entire 15 day period, the KEEP command, for example, was used by only two installations (one of which used it once), the RELEASE command was used only five times by one terminal, and nobody used the EXPLAIN command. Some of the commands are equipment-dependent (e.g., DISPLAY is not generally used without a CRT terminal and TYPE is recommended for use by all dialup terminals) so that the percentage distribution of command types reflects this factor also. The range of command use for the terminals is summarized in this table. One obvious difference is the relatively large number of TYPE commands used by the slow speed dialup printing terminals, and the proportionally large number of EXPAND commands used by the CRT display terminals. TYPE commands are generally used by non-CRT (hardcopy) terminals instead of DISPLAY. As expected, CRT terminals used many more DISPLAYS and hardcopy terminals (3, 88, 114, 125) used many more TYPES. The high speed terminals would also seem to encourage the use of the high data transfer commands such as EXPAND and DISPLAY.

Table 7 provide the same type of data as Table 6 except that the numbers and percentages are in terms of the time used by each of the commands. The completed time was the total elapsed time from the receipt of the command by the computer until the receipt of the next command. Table 7 is probably the most important data from the 15-day test records because the data is related directly to the terminal time used. Here we see several installations using considerably more of their terminal time for output functions (DISPLAY, TYPE, PAGE).

Four of the installations, all with mechanical, hardcopy non-CRT terminals, used approximately one quarter of their time on the TYPE command. It seems a possibility that some searchers use the terminal primarily to negotiate and arrange for a printout, while other searchers put more emphasis on the terminal being the actual output device. This point cannot be resolved without examining the nature and volume of printout requested by each of the terminals.

Another possible explanation for the large amount of time (proportionally) spent by the slow speed terminals on the output commands is that these commands involve a large amount of data transmission time. For example, a single TYPE command (assuming an average of 600 characters per item) would require 1.3 seconds on the 480 character per second terminal, and 20 seconds (and a correspondingly higher percentage of time used) on the 30 character per second terminal. Similarly an EXPAND command (assuming an average of 800 characters) would require 1.7 seconds on the 480 character per second terminal, and 27 seconds on the 30 character per second terminal. This is discussed further in a later section which explores the functional utilization of commands fcr each of the terminals.

Table 8 provides the data in terms of the number and percent of DIALOG commands used per question. This same data is plotted in Figure 13. The dominance of output or display commands for some installations can also be seen in this data. This table also shows the range of commands per

TABLE 6

COMMAND UTILIZATION PER TERMINAL (BY NUMBER OF COMMANDS)

	TERMINAL NUMBER												
	(SLOW TERMINAL EQUIPMENT)				(FAST TERMINAL EQUIPMENT								
NUMBER OF TIMES EACH COMMAND WAS USED	3	88	114	125	2	4	5	7	9	TOTAL	LOW	HIGH	HI/LO RATIO
SELECT	226 (33.93)*	168 (28.57)	1353 (44.43)	291 (28.64)	756 (30.20)	1217 (34.85)	1004 (28.93)	1521 (22.97)	897 (24.67)	7433 (29.69)	(22.97)	(44.43)	1.9
COMBINE	106 (15.92)	85 (14.46)	683 (22.43)	110 (10.83)	326 (13.02)	390 (11.17)	462 (13.31)	600 (9.06)	409 (11.25)	3171 (12.67)	(9.06)	(22.43)	2.5
EXPAND	5 (.75)	92 (15.65)	42 (1.38)	130 (12.80)	139 (5.55)	111 (3.18)	302 (8.70)	1661 (25.08)	568 (15.62)	3050 (12.18)	(.75)	(25.08)	33.4
PAGE	2 (.30)	13 (2.21)	3 (.10)	16 (1.57)	341 (13.62)	113 (3.24)	629 (18.13)	404 (6.10)	993 (27.32)	2514 (10.04)	(.10)	(27.32)	273.2
PRINT	91 (13.66)	41 (6.97)	170 (5.58)	70 (6.89)	305 (12.19)	422 (12.08)	202 (5.82)	529 (7.99)	216 (5.94)	2046 (8.17)	(5.58)	(13.66)	2.4
DISPLAY	0 (0)	0 (0)	0 (0)	0 (0)	174 (6.95)	354 (10.14)	395 (11.38)	673 (10.16)	257 (7.07)	1853 (7.40)	(0)	(11.38)	—
END	30 (4.51)	28 (4.76)	102 (3.35)	56 (5.51)	149 (5.95)	263 (7.53)	119 (3.43)	285 (4.30)	68 (1.87)	1100 (4.40)	(1.87)	(7.53)	4.0
TYPE	127 (19.07)	94 (15.99)	532 (17.47)	215 (21.16)	0 (0)	1 (.03)	55 (1.59)	20 (.30)	0 (0)	1044 (4.17)	(0)	(21.16)	—
LIMIT	29 (4.36)	27 (4.59)	30 (.99)	40 (3.94)	73 (2.92)	67 (1.92)	53 (1.53)	467 (7.05)	66 (1.82)	852 (3.40)	(.99)	(7.05)	7.1
BEGIN	22 (3.30)	31 (5.27)	111 (3.65)	49 (4.82)	43 (1.72)	187 (5.36)	78 (2.25)	144 (2.17)	65 (1.79)	730 (2.92)	(1.72)	(5.36)	3.1
DISPLAY SET HISTORY	7 (1.05)	6 (1.02)	0 (0)	15 (1.48)	63 (2.52)	229 (6.56)	65 (1.87)	79 (1.19)	0 (0)	464 (1.85)	(0)	(6.56)	—
RECALL	6 (.90)	0 (0)	0 (0)	4 (.39)	42 (1.68)	48 (1.37)	21 (.61)	87 (1.31)	26 (.72)	234 (.93)	(0)	(1.68)	—
EXECUTE	5 (.75)	0 (0)	0 (0)	5 (.49)	26 (1.04)	41 (1.17)	23 (.66)	82 (1.24)	26 (.72)	208 (.83)	(0)	(1.24)	—
SEND MESSAGE	2 (.30)	0 (0)	0 (0)	0 (0)	20 (.80)	21 (.60)	43 (1.24)	51 (.77)	31 (.85)	168 (.67)	(0)	(1.24)	—
FILE	8 (1.20)	3 (.51)	19 (.62)	15 (1.48)	2 (.08)	28 (.80)	19 (.55)	19 (.29)	13 (.36)	126 (.50)	(.08)	(1.48)	18.5
KEEP	0 (0)	0 (0)	0 (0)	0 (0)	39 (1.56)	0 (0)	0 (0)	1 (.02)	0 (0)	40 (.16)	(0)	(1.56)	—
RELEASE	0 (0)	0 (0)	0 (0)	0 (0)	5 (.20)	0 (0)	0 (0)	0 (0)	0 (0)	5 (.02)	(0)	(.20)	—
EXPLAIN	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)			
TOTAL	666 (100.00)	588 (100.00)	3045 (100.00)	1016 (100.00)	2503 (100.00)	3492 (99.88)	3470 (110.00)	6623 (100.00)	3635 (100.00)	25038 (100.00)			

* Figures in parentheses in the terminal number columns represent percent of commands used.

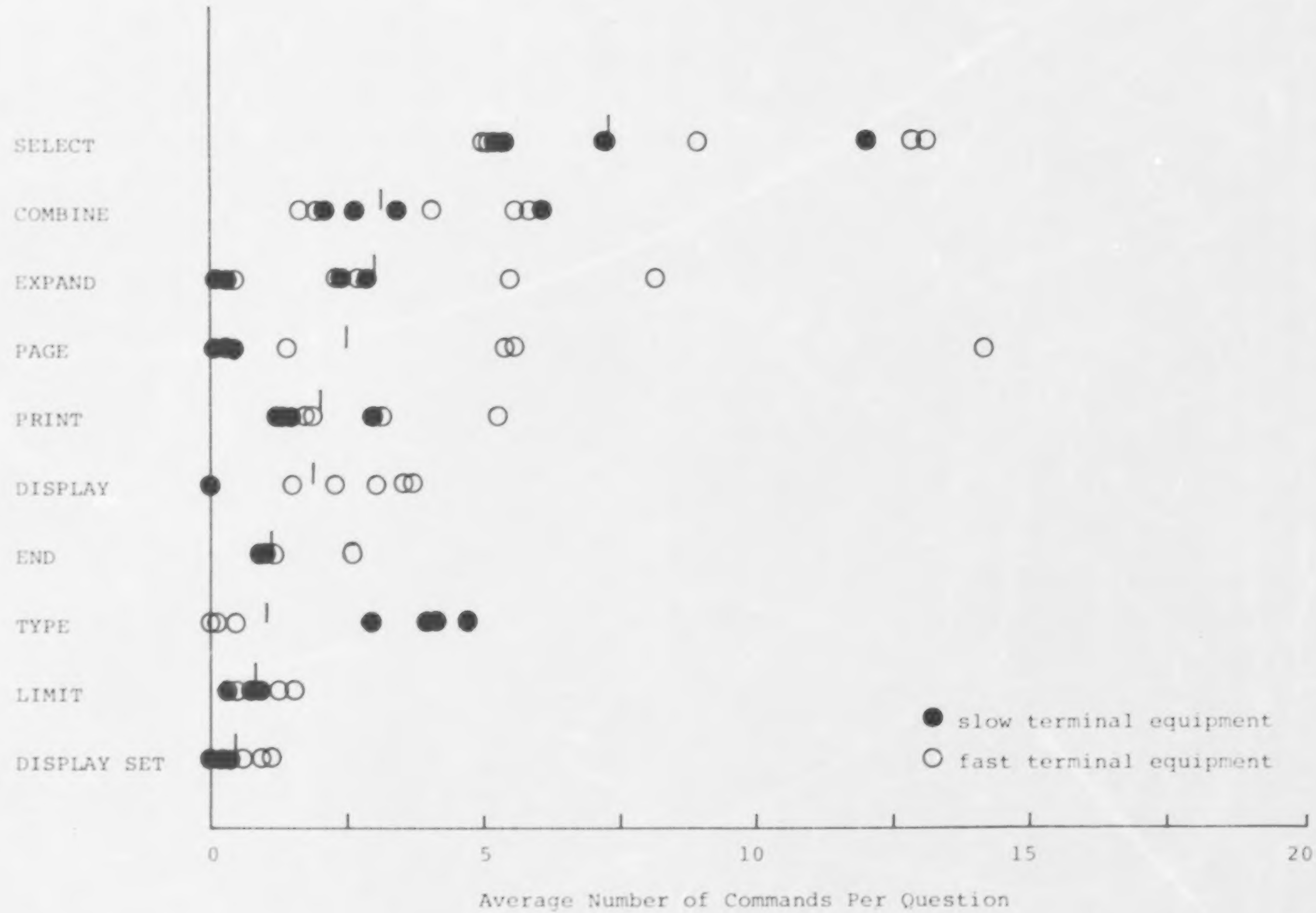


Fig. 13. Command Utilization Per Question

question searched by the installations. For the 15-day period there was an average of 24.76 commands per question for all terminals, although the individual terminal averages ranged from 14.60 to 51.92 commands per question.

Table 9 also provides data for each installation regarding command utilization per question. This data is in terms of terminal time used for each command. The average search time for this 15-day sample ranged from 5.5 to 29.0 minutes per question, with an average of 14.2 minutes per question for all searches done over this 15-day period.

Table 10 and Figure 14 provide data for each installation regarding command utilization per search, in terms of the number of commands used. Table 11 provides the same type of data expressed in terms of terminal time used.

Table 7 provides data for the total amount of terminal time used by each installation during this 15-day period (excluding the time associated with the END command). Table 6 provides data for the total number of commands used by each installation for this terminal time. This data assembled together in Table 12 provides another measure of the terminal activity at each installation (i.e., the rate at which commands are executed at the terminal). For the set of installations studied for the entire 15-day period, an average of about 85 commands per hour were entered at the slow speed hard copy terminals, and an average of about 112 commands per hour were entered at the fast terminal installations (with high speed CRT equipment).

9

TABLE 7

COMMAND UTILIZATION PER TERMINAL (BY TERMINAL TIME USED)

	TERMINAL NUMBER												
	(SLOW TERMINAL EQUIPMENT)				(FAST TERMINAL EQUIPMENT)								
TIME USED FOR EACH COMMAND (IN SEC.)	3	88	114	125	2	4	5	7	9	TOTAL	LOW	HIGH	HI/LOW RATIO
SELECT	5589 (20.58)	5388 (17.50)	35693 (31.13)	9905 (18.61)	20623 (20.46)	16957 (21.62)	27572 (19.43)	37244 (17.63)	9979 (9.77)	168950 (19.64)	(9.77)	(31.13)	3.2
BEGIN	5563 (20.49)	5489 (17.83)	17244 (15.04)	8732 (16.40)	13217 (13.11)	26134 (33.32)	18894 (13.31)	19355 (9.16)	9145 (8.95)	123773 (14.39)	(8.95)	(33.32)	3.7
EXPAND	226 (.83)	5594 (18.17)	2648 (2.31)	9304 (17.48)	6778 (6.73)	2431 (3.10)	18364 (12.94)	52543 (24.88)	22312 (21.84)	120200 (13.97)	(.83)	(24.88)	30.0
COMBINE	3721 (13.70)	2304 (7.48)	20793 (18.13)	4675 (8.78)	12917 (12.82)	8679 (11.06)	18831 (13.27)	15725 (7.45)	11412 (11.17)	99057 (11.51)	(7.45)	(18.13)	2.4
PAGE	203 (.75)	686 (2.23)	368 (.32)	897 (1.68)	13169 (13.07)	2205 (2.81)	18360 (12.94)	13803 (6.54)	24466 (23.95)	74157 (8.62)	(.32)	(23.95)	74.8
TYPE	7113 (26.19)	8815 (28.63)	32762 (28.57)	13500 (25.36)	0 (0)	1 (0)	3143 (2.21)	1121 (.53)	0 (0)	66455 (7.72)	(0)	(28.63)	---
DISPLAY	0 (0)	0 (0)	0 (0)	0 (0)	6431 (6.38)	6193 (7.89)	12499 (8.80)	24564 (11.63)	9364 (9.17)	59051 (6.86)	(0)	(11.63)	---
PRINT	1465 (5.39)	1194 (3.88)	3298 (2.88)	1631 (3.06)	7263 (7.21)	5472 (6.98)	8521 (6.00)	11457 (5.42)	3052 (2.99)	43353 (5.04)	(2.88)	(7.21)	2.5
LIMIT	1366 (5.03)	837 (2.72)	1022 (.89)	2007 (3.77)	3683 (3.65)	1355 (1.73)	2496 (1.76)	13621 (6.45)	1852 (1.81)	28239 (3.28)	(.89)	(6.45)	7.2
EXECUTE	644 (2.37)	0 (0)	0 (0)	518 (.97)	3856 (3.83)	2005 (2.56)	3090 (2.18)	9346 (4.43)	2582 (2.53)	22041 (2.56)	(0)	(4.43)	---
SEND MESSAGE	216 (.80)	0 (0)	0 (0)	0 (0)	810 (.80)	1248 (1.59)	5626 (3.96)	5983 (2.83)	6876 (6.73)	20759 (2.41)	(0)	(6.73)	---
DISPLAY SET HISITORY	453 (1.67)	386 (1.25)	0 (0)	1348 (2.53)	2949 (2.92)	4589 (5.85)	2164 (1.52)	2630 (1.24)	0 (0)	14519 (1.69)	(0)	(5.85)	---
RECALL	258 (.95)	0 (0)	0 (0)	159 (.30)	1686 (1.67)	602 (.77)	454 (.32)	3069 (1.45)	675 (.66)	6903 (.80)	(0)	(1.67)	---
RELEASE	0 (0)	0 (0)	0 (0)	0 (0)	6678 (6.62)	0 (0)	0 (0)	0 (0)	0 (0)	6678 (.78)	(0)	(6.62)	---
FILE	340 (1.25)	95 (.31)	835 (.73)	563 (1.06)	44 (.04)	568 (.72)	1924 (1.36)	746 (.35)	441 (.43)	5556 (.65)	(.04)	(1.36)	3.4
KEEP	0 (0)	0 (0)	0 (0)	0 (0)	694 (.69)	0 (0)	0 (0)	16 (.01)	0 (0)	710 (.08)	(0)	(.69)	---
EXPLAIN	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	(0)	(0)	---
TOTAL	27157 (100)	30788 (100)	114663 (100)	53239 (100)	100798 (100)	78439 (100)	141938 (100)	211233 (100)	102156 (100)	860401 (100)			

BEST COPY AVAILABLE

TABLE 8

COMMAND UTILIZATION PER QUESTION (BY NUMBER OF COMMANDS)

	TERMINAL NUMBER												
	(SLOW TERMINAL EQUIPMENT)				(FAST TERMINAL EQUIPMENT)								
NUMBER OF TIMES EACH COMMAND WAS USED	3	88	114	125	2	4	5	7	9	AVERAGE	LOW	HIGH	HI/LOW RATIO
SELECT	7.29 (33.92)	5.25 (28.56)	11.97 (44.43)	5.39 (28.62)	13.03 (30.20)	5.09 (34.86)	8.96 (28.92)	5.04 (22.97)	12.81 (24.67)	7.35 (29.68)	5.04 (22.97)	13.03 (44.43)	2.6 1.9
COMBINE	3.42 (15.91)	2.66 (14.47)	6.04 (22.42)	2.04 (10.83)	5.62 (13.02)	1.63 (11.16)	4.12 (13.30)	1.99 (9.07)	5.84 (11.25)	3.14 (12.68)	1.63 (9.07)	6.04 (22.42)	3.7 2.5
EXPAND	.16 (.75)	2.87 (15.61)	.37 (1.37)	2.41 (12.80)	2.40 (5.56)	.46 (3.15)	2.70 (8.72)	5.50 (25.07)	8.11 (15.62)	3.02 (12.20)	.16 (.75)	8.11 (25.07)	50.7 33.4
PAGE	.06 (.28)	.41 (2.23)	.03 (.11)	.30 (1.59)	5.88 (13.63)	.47 (3.22)	5.62 (18.14)	1.34 (6.11)	14.19 (27.33)	2.49 (10.06)	.03 (.11)	14.19 (27.33)	473 248.4
PRINT	2.94 (13.68)	1.28 (6.97)	1.50 (5.57)	1.30 (6.91)	5.26 (12.19)	1.77 (12.12)	1.80 (5.81)	1.75 (7.98)	3.09 (5.95)	2.02 (8.16)	1.28 (5.57)	5.26 (13.68)	4.1 2.4
DISPLAY	0 (0)	0 (0)	0 (0)	0 (0)	3.00 (6.95)	1.48 (10.14)	3.53 (11.39)	2.23 (10.16)	3.67 (7.07)	1.83 (7.39)	0 (0)	3.67 (11.39)	--- ---
END	.97 (4.51)	.88 (4.79)	.90 (3.34)	1.04 (5.52)	2.57 (5.96)	1.10 (7.53)	1.06 (3.42)	.94 (4.28)	.97 (1.87)	1.09 (4.40)	.88 (1.87)	2.57 (7.53)	2.9 4.0
TYPE	4.10 (19.08)	2.94 (16.00)	4.71 (17.48)	3.98 (21.14)	0 (0)	.00 (.00)	.49 (1.58)	.07 (.32)	0 (0)	1.03 (4.16)	0 (0)	4.71 (19.08)	--- ---
LIMIT	.94 (4.37)	.84 (4.57)	.27 (1.00)	.74 (3.93)	1.26 (2.92)	.28 (1.92)	.47 (1.52)	1.55 (7.07)	.94 (1.81)	.84 (3.39)	.27 (1.0)	1.55 (7.07)	5.7 7.1
BEGIN	.71 (3.30)	.97 (5.28)	.98 (3.65)	.91 (4.83)	.74 (1.71)	.78 (5.34)	.70 (2.26)	.48 (2.19)	.93 (1.79)	.72 (2.91)	.48 (1.71)	.98 (5.34)	2.0 3.1
DISPLAY SET HISTORY	.23 (1.07)	.19 (1.03)	0 (0)	.28 (1.49)	1.09 (2.53)	.96 (6.58)	.58 (1.87)	.26 (1.19)	0 (0)	.46 (1.86)	0 (0)	1.09 (6.58)	--- ---
RECALL	.19 (.89)	0 (0)	0 (0)	.07 (.37)	.72 (1.67)	.20 (1.37)	.19 (.61)	.29 (1.32)	.37 (.71)	.23 (.93)	0 (0)	.72 (1.67)	--- ---
EXECUTE	.16 (.75)	0 (0)	0 (0)	.09 (.48)	.45 (1.04)	.17 (1.17)	.21 (.68)	.27 (1.23)	.37 (.71)	.21 (.85)	0 (0)	.45 (1.23)	--- ---
SEND MESSAGE	.06 (.28)	0 (0)	0 (0)	0 (0)	.34 (.79)	.09 (.62)	.38 (1.23)	.17 (.77)	.44 (.85)	.17 (.69)	0 (0)	.44 (1.23)	--- ---
FILE	.26 (1.21)	.09 (.49)	.17 (.63)	.28 (1.49)	.03 (.07)	.12 (.82)	.17 (.55)	.06 (.27)	.19 (.37)	.12 (.48)	.03 (.07)	.28 (1.49)	9.3 21.3
KEEP	0 (0)	0 (0)	0 (0)	0 (0)	.67 (1.55)	0 (0)	0 (0)	.00 (.00)	0 (0)	.04 (.16)	0 (0)	.67 (1.55)	--- ---
RELEASE	0 (0)	0 (0)	0 (0)	0 (0)	.09 (.21)	0 (0)	0 (0)	0 (0)	0 (0)	.00 (.00)	0 (0)	.09 (.21)	--- ---
EXPLAIN	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	--- ---
TOTAL	21.49 (100.00)	18.38 (100.00)	26.94 (100.00)	18.83 (100.00)	43.15 (100.00)	14.60 (100.00)	30.98 (100.00)	21.94 (100.00)	51.92 (100.00)	24.76 (100.00)			

TABLE 9

COMMAND UTILIZATION PER QUESTION (BY TERMINAL TIME USED)

TIME USED FOR EACH COMMAND (IN SEC.)	TERMINAL NUMBER										LOW	HIGH	HI/LOW RATIO
	(SLOW TERMINAL EQUIPMENT)				(FAST TERMINAL EQUIPMENT)								
	3	88	114	125	2	4	5	7	9	AVERAGE			
SELECT	180.29 (20.58)*	168.37 (17.50)	315.87 (31.13)	183.43 (18.61)	355.57 (20.46)	70.95 (21.62)	246.18 (19.42)	123.32 (17.63)	142.56 (9.77)	167.11 (19.64)	(9.77)	(31.13)	3.2
BEGIN	179.45 (20.49)	171.53 (17.83)	152.60 (15.04)	161.70 (16.40)	227.88 (13.11)	109.35 (33.32)	168.70 (13.31)	64.09 (9.17)	130.64 (8.95)	122.43 (14.38)	(8.95)	(33.32)	3.7
EXPAND	7.29 (.83)	174.81 (18.17)	23.43 (2.31)	172.30 (17.48)	116.86 (6.73)	10.17 (3.10)	163.96 (12.94)	173.98 (24.87)	318.74 (21.84)	118.89 (13.97)	(.83)	(24.87)	30.0
COMBINE	120.03 (13.70)	72.00 (7.48)	184.01 (18.13)	86.57 (8.78)	222.71 (12.82)	36.31 (11.06)	168.13 (13.27)	52.07 (7.45)	163.03 (11.17)	.97.98 (11.52)	(7.45)	(18.13)	2.4
PAGE	6.55 (.75)	21.44 (2.23)	3.26 (.32)	16.61 (1.68)	227.05 (13.06)	9.23 (2.81)	163.93 (12.94)	45.71 (6.54)	349.51 (23.95)	73.35 (8.62)	(.32)	(23.95)	74.8
TYPE	229.45 (26.19)	275.47 (28.63)	289.93 (28.57)	250.00 (25.36)	0 (0)	0 (0)	28.06 (2.21)	3.71 (.53)	0 (0)	65.73 (7.72)	(0)	(28.63)	—
DISPLAY	0 (0)	0 (0)	0 (0)	0 (0)	110.88 (6.38)	25.91 (7.89)	111.60 (8.81)	81.34 (11.63)	133.77 (9.17)	58.41 (6.86)	(0)	(11.63)	—
PRINT	47.26 (5.40)	37.32 (3.88)	29.19 (2.88)	30.20 (3.06)	125.22 (7.21)	22.90 (6.98)	76.08 (6.00)	37.94 (5.42)	43.60 (2.99)	42.88 (5.04)	(2.88)	(7.21)	2.5
LIMIT	44.06 (5.03)	26.16 (2.72)	9.04 (.89)	37.17 (3.77)	63.50 (3.66)	5.67 (1.73)	22.29 (1.76)	45.10 (6.45)	26.46 (1.81)	27.93 (3.28)	(.89)	(6.45)	7.2
EXECUTE	20.77 (2.37)	0 (0)	0 (0)	9.59 (.97)	66.48 (3.82)	8.39 (2.56)	27.59 (2.18)	30.95 (4.43)	36.89 (2.53)	21.80 (2.56)	(0)	(4.43)	—
SEND MESSAGE	6.97 (.79)	0 (0)	0 (0)	0 (0)	13.97 (.80)	5.22 (1.59)	50.23 (3.96)	19.81 (2.83)	98.23 (6.73)	20.53 (2.41)	(0)	(6.73)	—
DISPLAY SET HISTORY	14.61 (1.67)	12.04 (1.25)	0 (0)	24.96 (2.53)	50.84 (2.92)	19.20 5.85)	19.32 (1.52)	8.71 (1.24)	0 (0)	14.36 (1.69)	(0)	(5.85)	—
RECALL	8.32 (.95)	0 (0)	0 (0)	2.94 (.30)	29.07 (1.67)	2.52 (.77)	4.05 (.32)	10.16 (1.45)	9.64 (.66)	6.83 (.80)	(0)	(1.67)	—
RELEASE	0 (0)	0 (0)	0 (0)	0 (0)	115.14 (6.63)	0 (0)	0 (0)	0 (0)	0 (0)	6.61 (.78)	(0)	(6.63)	—
FILE	10.97 (1.25)	2.97 (.31)	7.39 (.73)	10.43 (1.06)	.76 (.04)	2.38 (.72)	17.18 (1.36)	2.47 (.35)	6.30 (.43)	5.50 (.65)	(.04)	(1.36)	34.0
KEEP	0 (0)	0 (0)	0 (0)	0 (0)	11.97 (.69)	0 (0)	0 (0)	.05 (.01)	0 (0)	.70 (.08)	(0)	(.69)	—
EXPLAIN	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)			
TOTAL	876.02 (100.00)	[illegible] (100.00)	1014.72 (100.00)	985.90 (100.00)	1737.90 100.00)	328.20 (100.00)	1267.30 (100.00)	699.41 (100.00)	1459.37 (100.00)	851.04 (100.00)			

TOTAL TERMINAL TIME

* Figures in parentheses in the terminal number columns represent percent of terminal time used. The END time was excluded for all terminals for this analysis.

BEST COPY AVAILABLE

TABLE 10

COMMAND UTILIZATION PER SEARCH (BY NUMBER OF COMMANDS)

	TERMINAL NUMBER												
	(SLOW TERMINAL EQUIPMENT)				(FAST TERMINAL EQUIPMENT)								
NUMBER OF TIMES EACH COMMAND WAS USED	3	88	114	125	2	4	5	7	9	AVERAGE	LOW	HIGH	HI/LOW RATIO
SELECT	10.27 (33.93)*	5.42 (28.57)	12.19 (44.43)	5.94 (28.64)	17.58 (30.20)	6.51 (34.85)	12.87 (28.93)	10.56 (22.97)	13.80 (24.67)	10.18 (29.69)	5.42 (22.97)	17.58 (44.43)	3.2 1.9
COMBINE	4.82 (15.92)	2.74 (14.46)	6.15 (22.43)	2.24 (10.83)	7.58 (13.02)	2.09 (11.17)	5.92 (13.31)	4.17 (9.06)	6.29 (11.25)	4.34 (12.66)	2.09 (9.06)	7.58 (22.43)	3.6 2.5
EXPAND	.23 (.75)	2.97 (15.65)	.38 (1.38)	2.65 (12.80)	3.23 (5.55)	.59 (3.18)	3.87 (8.70)	11.53 (25.08)	8.74 (15.63)	4.18 (12.19)	.23 (.75)	11.53 (25.08)	50.1 33.4
PAGE	.09 (.30)	.42 (2.21)	.03 (.10)	.31 (1.57)	7.93 (13.62)	.60 (3.24)	8.06 (18.13)	2.81 (6.10)	15.28 (27.32)	3.44 (10.03)	.03 (.10)	15.28 (27.32)	509.3 273.2
PRINT	4.14 (13.67)	1.32 (6.97)	1.53 (5.58)	1.43 (6.89)	7.09 (12.19)	2.26 (12.08)	2.59 (5.82)	3.67 (7.99)	3.32 (5.94)	2.80 (8.17)	1.32 (5.58)	7.09 (13.67)	5.37 2.4
DISPLAY	0 (0)	0 (0)	0 (0)	0 (0)	4.05 (6.95)	1.89 (10.14)	5.06 (11.38)	4.67 (10.16)	3.95 (7.06)	2.54 (7.40)	0 (0)	5.06 (11.38)	--- ---
END	1.36 (4.50)	.90 (4.76)	.92 (3.35)	1.14 (5.51)	3.47 (5.95)	1.41 (7.53)	1.53 (3.43)	1.98 (4.30)	1.05 (1.87)	1.51 (4.39)	.90 (1.87)	3.47 (7.53)	3.8 4.0
TYPE	5.77 (19.07)	3.03 (15.99)	4.79 (17.47)	4.39 (21.16)	0 (0)	.01 (.03)	.71 (1.59)	.14 (.30)	0 (0)	1.43 (4.17)	0 (0)	5.77 (21.16)	--- ---
LIMIT	1.32 (4.36)	.87 (4.59)	.27 (.99)	.82 (3.94)	1.70 (2.92)	.36 (1.92)	.68 (1.53)	3.24 (7.05)	1.02 (1.82)	1.17 (3.40)	.27 (.99)	3.24 (7.05)	12.0 7.1
BEGIN	1.00 (3.30)	1.00 (5.27)	1.00 (3.65)	1.00 (4.82)	1.00 (1.72)	1.00 (5.36)	1.00 (2.25)	1.00 (2.17)	1.00 (1.79)	1.00 (2.92)	1.00 (1.72)	1.00 (5.36)	1.00 3.1
DISPLAY SET HISTORY	.32 (1.05)	.19 (1.02)	0 (0)	.31 (1.48)	1.47 (2.52)	1.22 (6.56)	.83 (1.87)	.55 (1.19)	0 (0)	.64 (1.85)	0 (0)	1.47 (6.56)	--- ---
RECALL	.27 (.90)	0 (0)	0 (0)	.08 (.39)	.98 (1.68)	.26 (1.37)	.27 (.61)	.60 (1.31)	.40 (.72)	.32 (.93)	0 (0)	.98 (1.68)	--- ---
EXECUTE	.23 (.75)	0 (0)	0 (0)	.10 (.49)	.60 (1.04)	.22 (1.17)	.29 (.66)	.57 (1.24)	.40 (.72)	.28 (.83)	0 (0)	.60 (1.24)	--- ---
SEND MESSAGE	.09 (.30)	0 (0)	0 (0)	0 (0)	.47 (.80)	.11 (.60)	.55 (1.24)	.35 (.77)	.48 (.85)	.23 (.67)	0 (0)	.55 (1.24)	--- ---
FILE	.36 (1.20)	.10 (.51)	.17 (.62)	.31 (1.48)	.05 (.08)	.15 (.80)	.24 (.55)	.13 (.29)	.20 (.36)	.17 (.50)	.05 (.08)	.36 (1.48)	7.2 18.5
KEEP	0 (0)	0 (0)	0 (0)	0 (0)	.91 (1.56)	0 (0)	0 (0)	.01 (.02)	0 (0)	.05 (.16)	0 (0)	.91 (1.56)	--- ---
RELEASE	0 (0)	0 (0)	0 (0)	0 (0)	.12 (.20)	0 (0)	0 (0)	0 (0)	0 (0)	.01 (.02)	0 (0)	.12 (.20)	--- ---
EXPLAIN	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)			
TOTAL	30.27 (100.00)	18.96 (100.00)	27.43 (100.00)	20.74 (100.00)	58.23 (100.00)	18.68 (100.00)	44.47 (100.00)	45.98 (100.00)	55.93 (100.00)	34.29 (100.00)			

* Figures in parentheses in the terminal number columns represent percent of commands used.

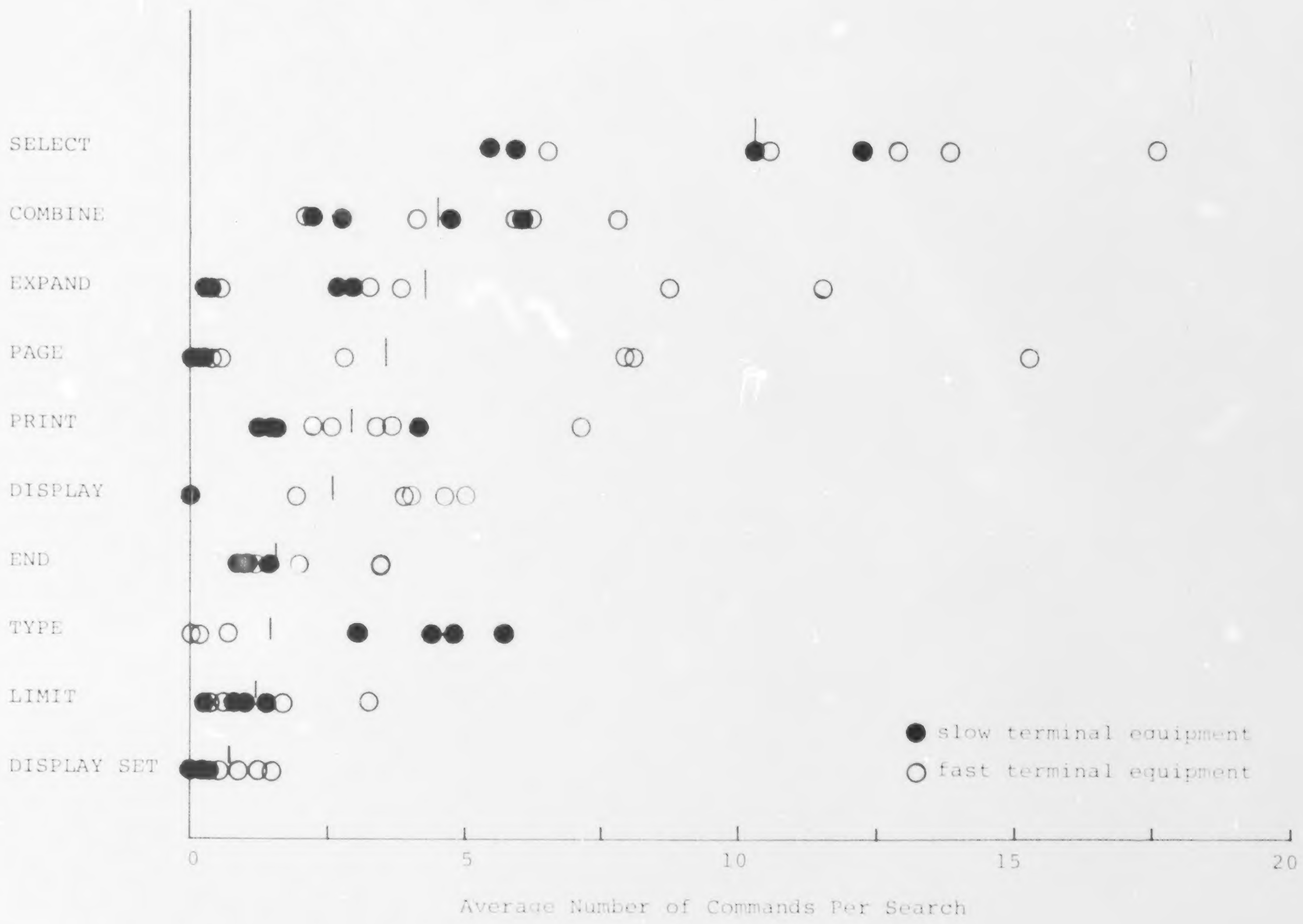


Fig. 14. Command Utilization Per Search

TABLE 11

COMMAND UTILIZATION PER SEARCH (BY TERMINAL TIME USED)

TIME USED FOR EACH COMMAND (IN SEC.)	TERMINAL NUMBER												
	(SLOW TERMINAL EQUIPMENT)				(FAST TERMINAL EQUIPMENT)								
	3	88	114	125	2	4	5	7	9	AVERAGE	LOW	HIGH	HI/LOW RATIO
SELECT	254.05 (20.58)*	173.81 (17.50)	321.56 (31.13)	202.14 (18.60)	479.60 (20.46)	90.68 (21.62)	353.49 (19.42)	258.64 (17.63)	153.52 (9.77)	231.44 (19.64)	(9.77)	(13.13)	3.2
BEGIN	252.86 (20.48)	177.06 (17.83)	155.35 (15.04)	178.20 (16.40)	307.37 (13.11)	139.75 (33.32)	242.23 (13.31)	134.41 (9.16)	140.69 (8.95)	169.55 (14.39)	(8.95)	(33.32)	3.7
EXPAND	10.27 (.83)	180.45 (18.17)	23.86 (2.31)	189.88 (17.48)	157.63 (6.72)	13.00 (3.10)	235.44 (12.94)	364.88 (24.88)	343.26 (21.84)	164.66 (13.97)	(.83)	(24.88)	30.0
COMBINE	169.14 (13.70)	74.32 (7.48)	187.32 (18.13)	95.41 (8.78)	300.40 (12.81)	46.41 (11.06)	241.42 (13.27)	109.10 (7.44)	175.57 (11.17)	135.69 (11.51)	(7.44)	(18.13)	2.4
PAGE	9.23 (.75)	22.13 (2.23)	3.32 (.32)	18.31 (1.69)	306.26 (13.07)	11.79 (2.81)	235.38 (12.94)	95.85 (6.53)	376.40 (23.95)	101.58 (8.62)	(.32)	(23.95)	74.8
TYPE	323.32 (26.20)	284.35 (28.63)	295.15 (28.57)	275.51 (25.35)	0 (0)	.01 (0)	40.29 (2.21)	7.78 (.53)	0 (0)	91.03 (7.72)	(0)	(28.63)	---
DISPLAY	0 (0)	0 (0)	0 (0)	0 (0)	149.56 (6.38)	33.12 (7.90)	160.24 (8.81)	170.58 (11.63)	144.06 (9.17)	80.89 (6.86)	(0)	(11.63)	---
PRINT	66.59 (5.39)	38.52 (3.88)	29.71 (2.88)	33.29 (3.07)	168.91 (7.21)	29.26 (6.97)	109.24 (6.00)	79.56 (5.43)	46.95 (2.99)	59.39 (5.04)	(2.88)	(7.21)	2.5
LIMIT	62.09 (5.03)	27.00 (2.72)	9.21 (.89)	40.96 (3.77)	85.65 (3.65)	7.25 (1.73)	32.00 (1.76)	94.59 (6.45)	28.49 (1.81)	38.68 (3.28)	(.89)	(6.45)	7.2
EXECUTE	29.27 (2.37)	0 (0)	0 (0)	10.57 (.97)	89.67 (3.83)	10.72 (2.56)	39.62 (2.18)	64.90 (4.43)	39.72 (2.53)	30.19 (2.56)	(0)	(4.42)	---
SEND MESSAGE	9.82 (.80)	0 (0)	0 (0)	0 (0)	18.84 (.80)	6.67 (1.59)	72.13 (3.96)	41.55 (2.83)	105.78 (6.73)	28.44 (2.41)	(0)	(6.73)	---
DISPLAY SET HISTORY	20.59 (1.67)	12.45 (1.25)	0 (0)	27.51 (2.53)	68.58 (2.93)	24.54 (5.85)	27.74 (1.52)	18.26 (1.25)	0 (0)	19.89 (1.69)	(0)	(5.85)	---
RECALL	11.73 (.95)	0 (0)	0 (0)	3.24 (.30)	39.21 (1.67)	3.22 (.77)	5.82 (.32)	21.31 (1.45)	10.38 (.66)	9.46 (.80)	(0)	(1.67)	---
RELEASE	0 (0)	0 (0)	0 (0)	0 (0)	155.30 (6.63)	0 (0)	0 (0)	0 (0)	0 (0)	9.15 (.78)	(0)	(6.63)	---
FILE	15.45 (1.25)	3.06 (.31)	7.52 (.73)	11.49 (1.06)	1.02 (.04)	3.04 (.72)	24.67 (1.36)	5.18 (.35)	6.78 (.43)	7.61 (.65)	(.04)	(1.36)	34.0
KEEP	0 (0)	0 (0)	0 (0)	0 (0)	16.14 (.69)	0 (0)	0 (0)	.11 (.01)	0 (0)	.97 (.08)	(0)	(.69)	---
EXPLAIN	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	(0)	(0)	---
TOTAL	1234.41 (100.00)	993.16 (100.00)	1033.00 (100.00)	1086.51 (100.00)	2344.14 (100.00)	419.46 (100.00)	1819.71 (100.00)	1466.82 (100.00)	1571.63 (100.00)	1178.64 (100.00)			

* Figures in parentheses in the terminal number columns represent present of terminal time used. The END time was excluded for all terminals for this analysis.

61

8

TABLE 12

AVERAGE COMMAND EXECUTION RATE FOR EACH TERMINAL INSTALLATION

Terminal	Total Number of Commands Entered	Total Number of Hours of Terminal Operation	Total Number of Questions	Average Number of Commands Entered Per Terminal Hour	Average Number of Questions Per Hour
SLOW PRINTER ONLY					
3	666	7.54	31	88.3	4.11
88	588	8.55	32	68.8	3.74
114	3,045	31.85	113	95.6	3.55
125	1,016	14.79	54	68.7	3.65
All Slow Terminals	5,315	62.73	230	84.7	3.67
FAST CRT WITH AUXILIARY PRINTER					
2	2,503	28.00	58	89.4	2.07
4	3,492	21.79	239	160.3	10.97
5	3,470	39.43	112	88.0	2.84
7	6,623	58.68	302	112.9	5.15
9	3,635	28.38	70	128.1	2.47
All Fast Terminals	19,723	178.28	781	111.9	4.43
ALL TERMINALS					
	25,038	239.01	1,011	104.8	4.23

D. CLASSIFICATION OF ERIC QUESTIONS

1. Classification Rules

In an attempt to explain differences in the rate of processing questions among the various installations, one suggested explanation was that perhaps some installations might be running a lot of "simple" questions while other installations are running "complex" and hence more time-consuming questions. To study this possibility, an algorithm was devised which classifies or grades queries submitted for on-line ERIC/DIALOG processing according to logical complexity. By applying the algorithm to all of the 1,011 questions submitted by all nine terminal installations during the 15 days of operation under investigation, a measure was obtained of the mix of question types submitted by a given terminal. It was felt that a comparison of these question mixes would be helpful in understanding why certain organizations processed more questions per hour than others.

The algorithm is intended to assign to any given question a rating of "simple", "moderate", or "complex" that is consistent with the judgment of logical complexity that might be made by persons experienced with automated information retrieval systems. Clearly many queries could reasonably be considered to be either of two neighboring categories. However, for the purposes of comparing general trends among several installations it was felt to be sufficient that the algorithm be consistent, and also assign a rating of logical complexity that agrees in a high percentage of cases with that of human judgment.

The classification algorithm takes into consideration several parameters of the search query. These are the total number of:

. DIALOG commands (N)

. SELECT commands (S)

. COMBINE commands (C)

. logical operators (L)

. AND operators (A).

Three different aspects of a query are considered by the algorithm: the total number of DIALOG commands, the number of SELECTs, and the apparent complexity of the query logic. A rating of "simple", "moderate", or "complex" is assigned independently for each of these three aspects of a question. If the same rating is assigned to two or three of these aspects, that rating becomes the rating for the entire question. If each of the three aspects is assigned a different rating, then the entire question is judged to be "moderate". All other possibilities are judged to be "simple".

The rating assigned to the first aspect (total number of DIALOG commands) is made as follows. The total number of DIALOG commands is counted, but in so doing only 1/3 of the DISPLAY, TYPE, PRINT, and PAGE commands are counted. These commands are given less weight because they represent output functions rather than search strategy functions. The resulting "N" total determines the command rating as follows:

N	Rating
< 15	simple
15-30	moderate
> 30	complex.

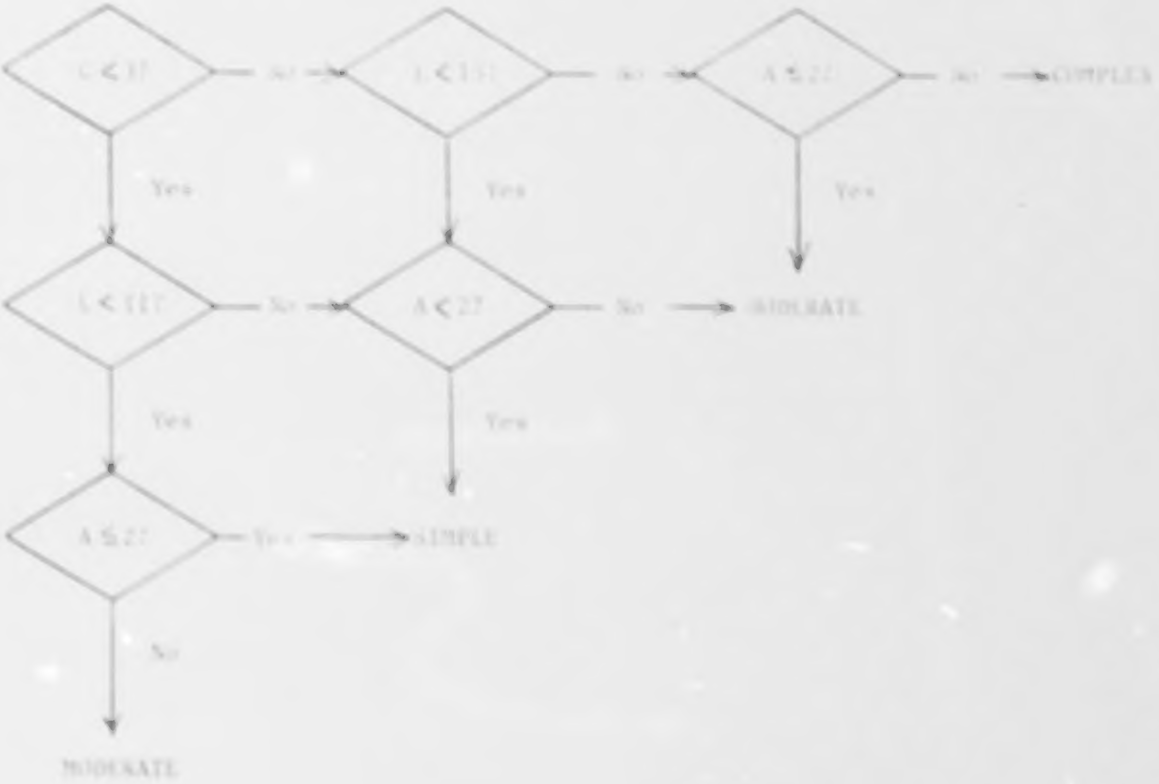
The rating assigned to the second aspect (total number of SELECTS) is made by simply counting the number of SELECT commands, "S", in the question and applying this rule:

S	Rating
< 9	simple
9-14	moderate
> 14	complex.

The rating assigned to the third and final aspect (complexity of search logic) is made by applying the rules indicated below. ("A" is the number of AND operators, "C" the number of COMBINE commands, "L" the number of logical operators, "S" the number of SELECTs.)

1. If $A \geq 3$, and $S > 12$, then the question is judged to be complex.

2. Otherwise apply the following decision chart.



In evaluating and calibrating this algorithm, two staff members made independent "simple", "moderate", or "complex" judgments of approximately 125 of the questions processed during the 15 days under study. The ratings assigned by the algorithm to these questions agreed with the composite judgment of the staff members as consistently (85-90%) as the individual staff members agreed among themselves. This suggests that the algorithm provides a consistent rating approximately equivalent to that which would be obtained by a manual examination of each question.

As a point for possible improvement of the algorithm, it was noted after all of the work had been done that we had underestimated the number of OR operations in some of the questions. We unfortunately did not examine and compute an equivalent number of OR operations for those searches that used a SELECT range. Our algorithm did not recognize SELECT E1-E6,E8,E10 as implying 7 ORs instead of none. The terminal that used such a composite command was undercounted in the number of SELECT commands and ORs compared to those that would have been counted if the searcher had SELECTed terms individually and then COMBINEd them later. In these cases the number of SELECT commands does not correspond to the number of descriptors used. (Another case is the use of SEARCH SAVEs, which commonly contain many descriptors and appropriate logical operators, retrieved as one set.)

2. The Data

The results of applying this algorithm to all questions submitted by nine terminals during the 15 days under study are shown in Figure 15. It can be seen that there is a considerable difference in the mix of question complexity associated with each of the installations. One installation had a high of about 35% complex questions, while another installation had less than 4% complex questions. Several aspects of this issue are discussed below.

a. Search Time as a Function of Question Complexity

Intuitively, one would expect that complex questions would take more on-line time than simple questions. The data from this study tends to supports that safe hunch. Figure 16 shows on a terminal-by-terminal basis, how the average search rate correlated with the percent of simple questions processed by that terminal. The percentage of simple questions processed by each terminal was taken from the data in this section. The search rates used in this figure were the rates experienced for these terminals during the same general period (October-November 1973) that this question complexity data was drawn from. It can be seen from this figure that there is some slight correlation between these two factors, but not as pronounced as one might expect.

b. Question Complexity as a Function of the Installation

One might suggest that the question mix might be influenced by the particular type of installation, sponsoring organization, constituency, or user group that is being served by the terminal installation. Unfortunately, this study did not collect any data that could be used to investigate this question. We do know that the installations were serving different types of user groups.

c. Question Complexity as a Function of the Terminal Equipment

It seemed possible that the question mix might be related to the type of terminal equipment, for the reason that a terminal operator might be more inclined to use more EXPAND and DISPLAY commands if they could be swiftly executed. Figure 16 provides some data on this point, and shows that there does not appear to be any strong correlation on this point.

d. Question Mix as a Function of Personal Work Habits of the Analysts

In the work of classification and indexing, it has been known for years that there are differences in the approach and results when two or more people do the same task. Even the same indexer repeating a given task at a later date may be inconsistent in the assignment of indexing terms. The reports of many indexer consistency tests have made this point. We now have an analogous situation in which it seems quite likely that two or more profile analysts or terminal operators, given the same information request and conditions, will generate different search statements. It also seems possible that given the same information request, one person could

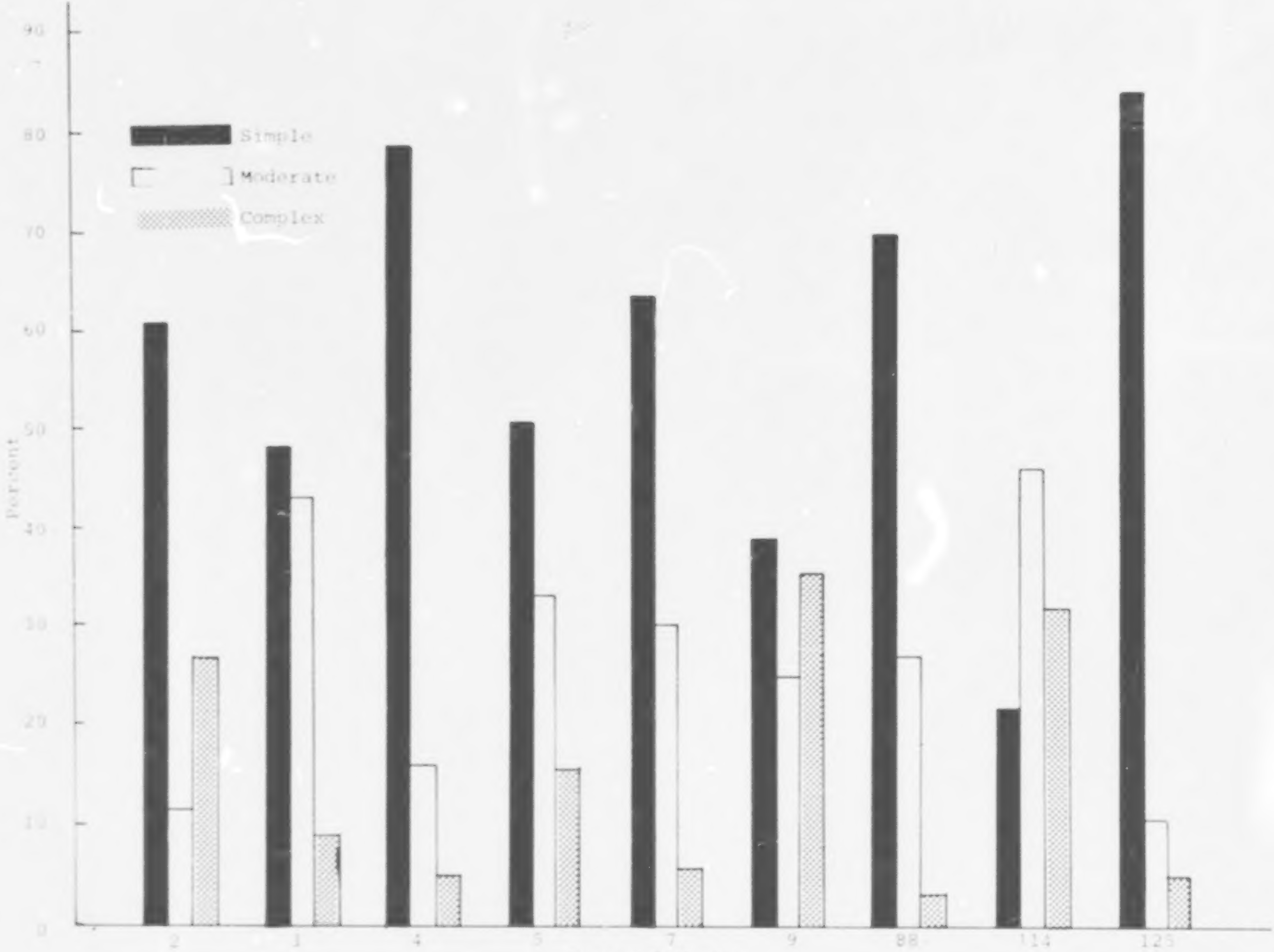


Fig. 15. Mix of Question Complexity for Each Terminal

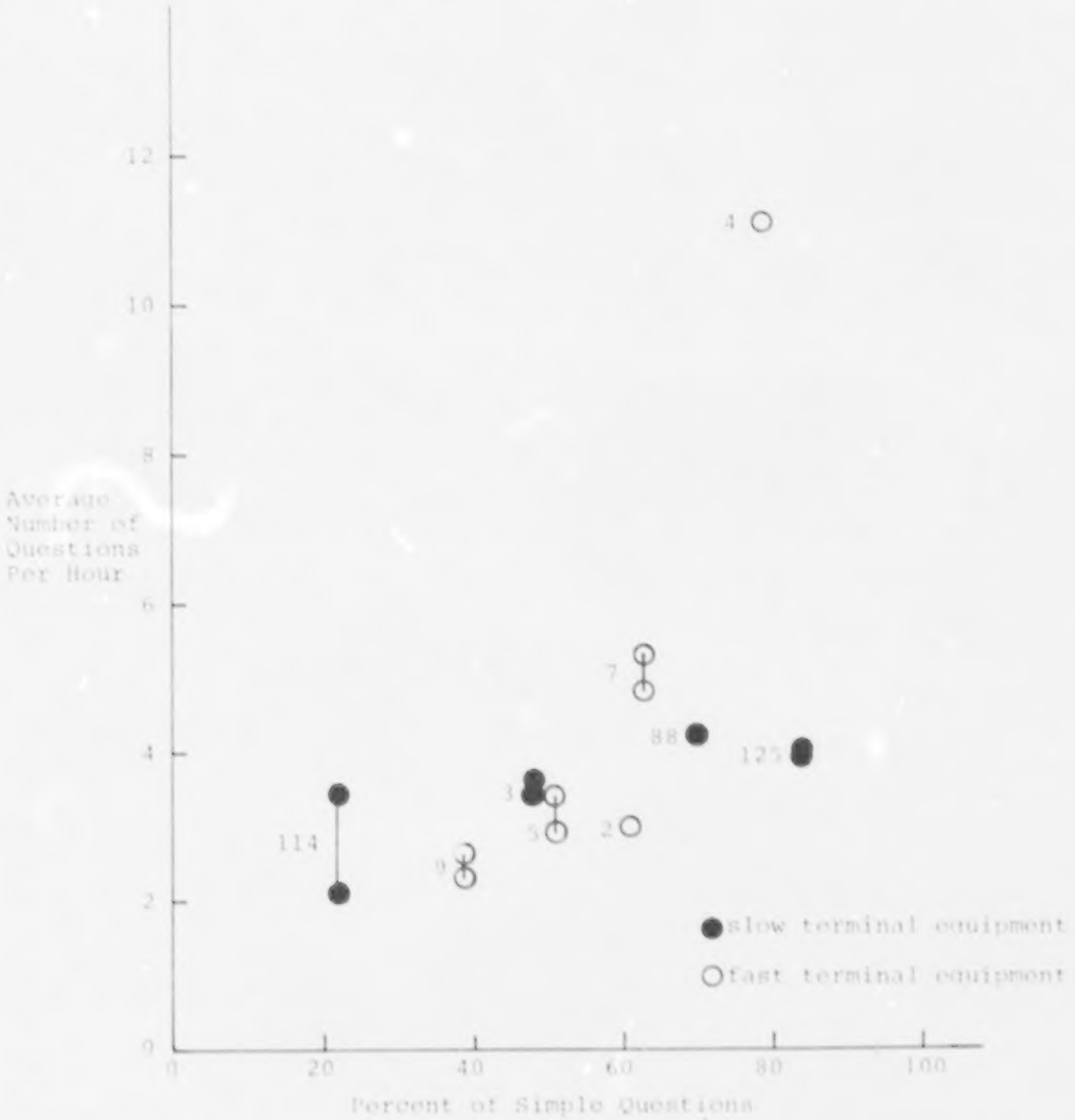


Fig. 16. Relationship of Search Rates to Search Complexity

come up with a simple search while another person came up with a complex search (and they might even get approximately the same results). We have seen instances in which this has happened. We have talked to analysts who readily agree that they always try to make their search as comprehensive as possible, and we have also talked to analysts who make a point of always trying to make a search as simple as possible in order to extract a few of the most relevant citations.

Thus we see a possible pattern in which question complexity is a matter of personal style and work habits, or personal approach to a problem (or perhaps maybe even a matter of institutional style or policy). This is a significant factor for this study because most of the terminal installations use only a few terminal operators (typically the bulk of searching is done by 1-3 different individuals at each site), consequently, the pattern of a single operator can in fact be the pattern for the installation.

No data was collected during this study that could be used to support or reject this hypothesis. However, this point of view was confirmed in many of our discussions with terminal operators from other installations. The data from the recent BIOSIS test (26 terminal installations running the same 20 questions) should provide some very good information on this point.

E. VARIANT FORMS OF ERIC DESCRIPTORS AND IDENTIFIERS

1. Background

The EXPAND command results in a display such as that shown in Figure 17 that is a mixture of both Descriptors and Identifiers. The 7,520 Descriptors are controlled terms from the ERIC Thesaurus, have a rigid authority control procedure associated with their input, and are seldom in error. The Identifiers are not subjected to the same quality control and review procedures as the Descriptors, and this has resulted in a considerable degree of inconsistency and error. However, efforts are underway to standardize use of Identifiers.

Because of these differences and for other reasons, the Descriptor and Identifier files were at one time separated in the DIALOG system. However, for at least the last year they have been combined in a single file so that a mixed collection of Descriptors and Identifiers are displayed as a result of the EXPAND command.

During our early use of the DIALOG system we became aware of many instances in which the EXPAND command would show one or more variant forms of the same word (e.g., both the singular and plural forms of the same word). Because the variant forms occurred so frequently, it was felt that perhaps some conscientious searchers would anticipate their occurrence and would use more EXPAND and SELECT commands than other searchers, and that this might contribute to an increase in the average length of the searches, if done consistently. We were also concerned with the retrieval loss that might be experienced by not including in the search statement all of the variant forms of words. For these reasons, we decided to explore in more detail the nature and frequency of the variant forms, and their impact on the search process and results.

2. Nature and Extent of Variant Forms

An analysis of the printed ERIC term posting frequencies would have provided some useful information about the frequency of occurrence of variant forms, but would not have led directly to information about their impact on searching. For that reason, it was decided instead to study a number of representative real searches that had been done by other installations. Using the command histories provided by Lockheed for the nine terminals under study, a total of 80 searches were sampled randomly from three days of DIALOG operation. As described earlier, a search was defined as any one discrete command history with a BEGIN and END command, and usually consisted of one or more EXPAND, SELECT, and COMBINE commands with or without a PRINT command.

Each term that was SELECTed by the searcher in these 80 searches (e.g., IT=AID, IT=SUMMER SCHOOL) was looked up in the July 1973 issue of ERIC/DIALOG Cumulative Listing of Descriptor and Identifier Usage in RIE and CIJE to see if there were any variant forms of this term, and to see to what extent the searchers picked up the variant forms.

```
      EXPAND EPS=CALIFORNIA UNIV., BERKELEY. INST. OF LIBRARY RESEARCH
REF   INDEX-TERM             TYPE ITEMS RT
E1    PS=CALIFORNIA UNIV.,
        BERKELEY. INST. OF LI-      32
E2    PS=CALIFORNIA UNIV.,
        BERKELEY. INST. OF PE-       3
E3    PS=CALIFORNIA UNIV.,
        BERKELEY. INST. OF UR-       2
E4    PS=CALIFORNIA UNIV.,
        BERKELEY. JAPANESE LI-      12


                 EPS=CALIFORNIA UNIV., LOS ANGELES. INST. OF LIBRARY RESEARC

REF   INDEX-TERM             TYPE ITEMS RT
E1    PS=CALIFORNIA UNIV.,
        LOS ANGELES. DEPT. OF-       7
E2    PS=CALIFORNIA UNIV.,
        LOS ANGELES. DIV. OF--      56
E3    PS=CALIFORNIA UNIV.,
        LOS ANGELES. EARLY CH-      22
E4    PS=CALIFORNIA UNIV.,
        LOS ANGELES. ERIC CLE-     204
E5    PS=CALIFORNIA UNIV.,
        LOS ANGELES. GRADUATE-      30
E6    PS=CALIFORNIA UNIV.,
        LOS ANGELES. INST. OF-       5
E7    PS=CALIFORNIA UNIV.,
        LOS ANGELES. INST. OF-      11
E8    PS=CALIFORNIA UNIV.,
        LOS ANGELES. INST. OF-      49
E9    PS=CALIFORNIA UNIV.,
        LOS ANGELES. INST. OF-       1
E10   PS=CALIFORNIA UNIV.,
        LOS ANGELES. JUNIOR C-      10
                                   -MORE-
```

which Inst. ? If complete, correct version
had not been entered, one would
not be able to tell

BEST COPY AVAILABLE

Fig. 17a. EXPAND Showing Corporate Authors (PS or Publication Source Field)

```
                    ECALIFORNIA UNIVERSITY (LOS ANGELES)
REF  INDEX-TERM                TYPE ITEMS RT
E1   IT=CALIFORNIA TEST OF
       PERSONALITY (CTP)-----       3
E2   IT=CALIFORNIA UNIVERSITY      12
E3   IT=CALIFORNIA
       UNIVERSITY (BERKELEY)-       4
E4   IT=CALIFORNIA
       UNIVERSITY (DAVIS)----       1
E5   IT=CALIFORNIA
       UNIVERSITY (IRVINE)---       1
E6  -IT=CALIFORNIA UNIVERSITY
        (LOS ANGELES)--------       2
E7   IT=CALIFORNIA UNIVERSITY
        (SAN DIEGO CENTRA----       1
E8   IT=CALIFORNIA UNIVERSITY
        (SAN DIEGO)----------       2
E9   IT=CALIFORNIA UNIVERSITY
        (SAN FRANCISCO)------       1
E10  IT=CALIFORNIA UNIVERSITY
        AT BERKELEY----------       1
                                  -MORE-
                   PAGE
REF  INDEX-TERM                TYPE ITEMS RT
E11  IT=CALIFORNIA UNIVERSITY
        AT LOS ANGELES-------       1
E12  IT=CALIFORNIA UNIVERSITY
        AT SAN DIEGO---------       3
E13  IT=CALIFORNIA UNIVERSITY
        AT SANTA BARBARA-----       1
E14  IT=CALIFORNIA UNIVERSITY
        AT SANTA CRUZ--------       1
E15  IT=CALIFORNIA
       UNIVERSITY BERKELEY---       9
E16  IT=CALIFORNIA UNIVERSITY
        MEDICAL CTR (SAN-----       1

                    EUNIVERSITY OF CALIFORNIA
REF  INDEX-TERM                TYPE ITEMS RT
E1   IT=UNIVERSITY OF
       BRIDGEPORT------------       3
E2   IT=UNIVERSITY OF
       BRITISH COLUMBIA------       7
E3   IT=UNIVERSITY OF BUFFALO       2
E4   IT=UNIVERSITY OF
       BUFFALO (NEW YORK)----       1
E5   IT=UNIVERSITY OF CALGARY       2
E6  -IT=UNIVERSITY OF
       CALIFORNIA------------      69
E7   IT=UNIVERSITY OF
       CALIFORNIA (BERKELEY)-       7
E8   IT=UNIVERSITY OF
       CALIFORNIA (IRVINE)---       1
E9   IT=UNIVERSITY OF CALIFOR
       NIA (LOS ANGELES)-----      11
```

BEST COPY AVAILABLE

Fig. 17b. EXPAND Showing Corporate Authors (Identifiers)

Variant forms were identified in this sample as being of the following types:

-- MORPHOLOGICAL

. singular vs. plural form (TEST, TESTS)

. gerund form (TESTING)

. possessive forms (BLOOM TAXONOMY, BLOOMS TAXONOMY)

-- SPELLING FORM

. English vs. American forms (LABOUR, LABOR)

. Acronyms (IGE, INDIVIDUALLY GUIDED EDUCATION)

. Abbreviations (CAL., CALIF., CALIFORNIA)

. Compound nouns with or without space or hyphen (POST SECONDARY EDUCATION, POST-SECONDARY EDUCATION, POSTSECONDARY EDUCATION; FILM STRIP, FILMSTRIP)

-- SPELLING ERRORS (COUNSELING GOALS, COUNSELING FOALS).

One extreme example of variant entries is Title III of the Elementary and Secondary Education Act which is listed in 17 different ways. Our analysis work also considered some words or variant forms that are often used in a synonymous way (e.g., CHICANO, MEXICAN-AMERICAN).

Our analysis of all of the 80 command histories resulted in the data shown in Table 13. Several observations can be made regarding the data in this table:

. There are many variant forms in this data base. Comprehensive searching would have included 123 variant forms in a total of 764 terms (16.1%). About one out of every 6 terms had a variant form which could have been added to the EXPAND and/or SELECT operation if the searcher desired.

. The searchers did not use many variant forms in their searches. Consequently, this probably did not significantly influence the time required to do the searches. Furthermore, the searcher's use of variant forms was distributed rather evenly over all the terminals so that the inclusion of variant forms in the search statement probably did not contribute significantly to the difference in average search speeds between the terminals.

A more detailed analysis of the variant forms encountered in this study led to the data shown in Table 14. The data from Tables 13 and 14 suggest a need for some quality control improvements in the data base. However, we did check further to see what impact the use or non-use of these variant forms would have on the search results.

TABLE 13

EXTENT OF VARIANT FORMS USED OR NOT USED IN 80 REAL SEARCHES

	Oct. 30 Searches	Oct. 31 Searches	Nov. 1 Searches	All Searches
Total number of searches examined	21	39	20	80
Total number of terms originally used in these searches (i.e. SELECT)	176	370	114	660
Total number of variant terms actually used by the original searchers	9	10	0	19
Total number of additional variant terms found by our lookups, that could also have been used by the original searchers but were not	30	55	19	104
Total number of variant forms that could have been used	39	65	19	123

TABLE 14

CHARACTERISTICS OF VARIANT FORMS USED OR NOT USED IN 80 REAL SEARCHES

Types of Variant Forms	Number of Occurrences	Number of Items
Singular	34	69
Plural	15	191
Gerund	1	11
Possessive	2	2
Spelling	3	5
Spacing and hyphenation	12	116
Acronyms and abbreviations	5	8
Errors	19	42

3. Impact of Variant Forms on Search Output

For this part of the study, nine of the previous 80 searches were chosen for further analysis because they frequently used descriptors that had variant forms. In order to see what effect the variant forms had on the search results, the search steps of the original searches were re-created, followed by another search that incorporated all of the possible variant forms in the place where they would have been used in the original search. The number of output citations was noted after each COMBINE operation, in both the original and augmented searches. No attempt was made to judge the relevance of the selected citations. The results of these searches are shown in Table 15.

The data from this study seem to indicate that although there are a significant number of variant forms of subject terms that could be incorporated into the searches, the addition of these variant forms to the searches does not significantly affect the search results in terms of retrieving a a large number of additional citations. In seven of the nine sample searches, the results stayed the same when a total of 28 variant forms was added to the original searches. In the remaining two sample searches (searches 8,9 in Table 15) a total of seven additional citations was added to the original 450 citations as a result of the inclusion of 12 additional variant forms, for an increase of about 1.5 percent of the original citations for those two searches.

As guidelines for the searchers, the data would suggest that if the most important Descriptors and Identifiers were used in the search, the redundancy of indexing is such that the lookup and inclusion of every possible variant form of descriptor or identifier may not be necessary unless the highest possible recall is an objective of the search. One major exception to this practice is the handling of variant forms of author names and institution names. There are many variant forms for these names, and they should be EXPANDed and included in all variant forms.

It is planned that in January 1975 a complete revision of the ERIC/DIALOG data base will be made available. The new data base will be offered with the same powerful full-text indexing techniques currently available on all other DIALOG files. Full-text indexing will include the title, Descriptor, Identifier and corporate author fields. The searcher will be able to retrieve the bound Descriptor and Identifier phrases as done now, but in addition, the searcher will be able to locate any word pattern, including word distance and order, contained in any combination of the full-text indexed fields. Full-text SELECT operations allow the specification of inter-word distances at the word, sentence, field or citation levels in any combination. This facility will greatly simplify the process of collecting word form variations as well as synonyms. For example, by SELECTing the term READING/DE,ID the searcher will immediately obtain all uses of the word READING in any Descriptor or Identifier regardless of its word position. Thus postings to hundreds of ERIC Descriptors and Identifiers will have been precombined for the searcher.

TABLE 15

Comparison of Original Search Results With the Results Obtained By

Using All Possible Variant Forms of Descriptors and Identifiers

Search Number	Terminal Number	Number of Variant Forms That Could Have Been Added to The Original Search	Total Number of Citations Posted Under The Additional Variant Forms	Number of Output Citations Without/With The Additional Variant Forms
1	2	5	5	0/0; 1/1; 0/0
2	2	4	6	91/91; 15/15
3	2	5	6	0/0
4	4	2	4	11/11
5	125	2	2	13/13
6	125	4	13	51/51
7	6	6	700	91/91
8	5	2	106	47/51; 6/6
9	2	10	14	90/93; 99/99; 59/59; 149/149

F. INFLUENCE OF OTHER FACTORS ON SEARCH SPEEDS

1. Terminal Equipment

In theory, the type and speed of terminal equipment definitely influences the search strategy, the command utilization, and the productivity of the terminal installation; that relationship is not clearly borne out in the test data. The data in Table 12 clearly shows that for the 15 day detailed sample, the fast CRT displays, as a group, execute about 1.3 times as many commands per terminal hour than the slow speed terminals do, and about 1.2 times as many questions per hour . However over the entire span of terminal operation described in Table 1 and Figure 1, a mixed trend is seen -the slow speed mechanical terminals, as a group, seemed to run more searches per hour than the fast CRT terminals, especially for the last third of the period that is shown in Figure 1. It would seem that the data does show that the type and speed of the terminal equipment is in fact a significant factor in explaining some but not all of the differences in search productivity for the installations studied.

Table 16 does show that there are clear differences in command utilization by terminal type. Both types of terminals used about the same percent of their commands for query formulation and negotiation (in the range of about 54-69% of all commands used). However the slow mechanical terminals used a greater percent of their commands for output functions (about 23-33%) than was used by the fast CRT terminal installations (about 13-22%).

A greater percent of TYPE commands was used with the slow terminals, in comparison to the equivalent DISPLAY command for the fast CRT terminals. Probably this was because the slow terminals had hard copy output as a result of search negotiation operations that could also be used for immediate search results, particularly for searches resulting in a small number of citations.

2. Continuing Education, and Association with other Searchers

It seems quite possible that a searcher who was isolated from other searchers would not continue to develop the searching skills and performance that might otherwise be possible. A searcher working with a large group of other searchers within the same institution, would be in a position to share ideas and techniques to gradually upgrade the performance of the entire group of searchers. Similarly, participation in user groups, continuing training by representatives of the on-line service, and site visits to other terminal installations, would all seem to be positive influences in upgrading searcher performance. It is quite possible that some of the installations included in this study operated with a very small staff of searchers (e.g. 1-3), and were relatively limited in the extent to which they could take advantage of these opportunities for continuing education and training. This factor might explain some of the differences in installation productivity.

TABLE 16

FUNCTIONAL UTILIZATION OF COMMANDS (BY PERCENT OF COMMANDS USED)

COMMAND FUNCTION	TERMINAL NUMBER								
	(SLOW TERMINAL EQUIPMENT)				(FAST TERMINAL EQUIPMENT)				
	3	88	114	125	2	4	5	7	9
SYSTEM CONTROL									
BEGIN	3.30	5.27	3.65	4.82	1.72	5.36	2.25	2.17	1.79
END	4.51	4.76	3.35	5.51	5.95	7.53	3.43	4.30	1.87
PAGE	.30	2.21	.10	1.57	13.62	3.24	18.13	6.10	27.32
DISPLAY SET HISTORY	1.05	1.02	0	1.48	2.52	6.56	1.87	1.19	0
SEND MESSAGE	.30	0	0	0	.80	.60	1.24	.77	.85
FILE	1.20	.51	.62	1.48	.08	.80	.55	.27	.36
KEEP	0	0	0	0	1.56	0	0	.02	0
RELEASE	0	0	0	0	.20	0	0	0	0
EXPLAIN	0	0	0	0	0	0	0	0	0
	10.66	13.77	7.72	14.86	26.45	24.09	27.47	14.84	32.19
QUERY FORMULATION & NEGOTIATION									
SELECT	33.93	28.57	44.43	28.64	30.20	34.85	28.93	22.97	24.67
EXPAND	.75	15.65	1.38	12.80	5.55	3.18	8.70	25.08	15.62
COMBINE	15.92	14.46	22.43	10.83	13.02	11.17	13.31	9.06	11.25
EXECUTE	.75	0	0	.49	1.04	1.17	.66	1.24	.72
RECALL	.90	0	0	.39	1.68	1.37	.61	1.31	.72
LIMIT	4.36	4.59	.99	3.94	2.92	1.92	1.53	7.05	1.82
	56.61	63.27	69.23	57.09	54.41	53.66	53.74	66.71	54.80
OUTPUT									
PRINT	13.66	6.97	5.58	6.89	12.19	12.08	5.82	7.99	5.94
DISPLAY	0	0	0	0	6.95	10.14	11.38	10.16	7.07
TYPE	19.07	15.99	17.47	21.16	0	.03	1.59	.30	0
	32.73	22.96	23.05	28.05	19.14	22.25	18.79	18.45	13.01
TOTAL	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00

The DIALOG Users Group met for the first time in early 1973. The ERIC Users Group also met for the first time in 1973. It is not clear what type of person attended these meetings (e.g. managers instead of searchers), and it might be that these meetings did not contribute significantly to the terminal performance that we measured for 1973. During this same time period, DIALOG representatives were visiting each of the terminal installations and answering questions, but did not have a formal program of continuing education. The instruction manuals, newsletters, and other documentation materials were not as well developed as they are today, and may not have been a factor in improving the terminal performance in 1972-73. However, it should be noted that many of the installations have shown a continually improving performance picture as the documentation and user communication channels improved.

What we can say about these communication factors is that we feel that they can influence terminal productivity, however we have no direct data from this study to support that feeling.

3. Subject Expertise

It would seem reasonable to expect higher performance from searchers who were subject specialists in the topics being searched. One would also expect that searchers at the ERIC clearinghouses would be particularly proficient because they knew the data base and the indexing terminology. One of our test installations that was an ERIC clearinghouse did in fact have a high search rate. No data was available however to relate the individual searchers and their backgrounds, to the searches analyzed during this study.

4. Extent of Pre-Planning Before Searching

Almost all of the installations followed the practice of doing some preparatory work before searching at the terminal. This is clearly seen by most installations as a practice which can result in more effective use of terminal time. Some installations insist on this approach as part of their operating policy and procedures. One of the test installations that did not follow this practice did have a relatively low search rate.

If this practice is followed too closely in the quest for increased on-line productivity, with little discretionary work at the terminal, it is possible that the whole character of on-line searching can be changed from an interactive dialog to a remote-job entry situation. This would be unfortunate because it would deny us some of the important advantages of interactive searching. There is clearly a tradeoff between additional preparatory time and time spent on-line. A rational approach to productivity enhancement will try to minimize total cost.

5. Fee Versus Free Service, and Cost-Conscious Attitudes

The cost-conscious attitude of the searchers or their institution seemed to have an important influence on the terminal productivity. Searchers who were operating in an environment in which the searching costs were fully subsidized and were perceived as "free" by the searcher, used the terminal in a different manner than those searchers who were operating in a cost-recovery or full charging mode. The searcher who visualizes a taxi cab meter mounted on the side of the terminal and ticking off dollars to correspond to terminal time, is much more anxious to get the search completed as soon as possible. This attitude has been confirmed in many discussions with searchers and installation managers, both for the installations in this study and elsewhere.

Table 17, restructuring the data from Table 12, provides a summary of the performance data for the installations included in this study. The charging services run about 1.3 times more commands per hour through their terminals than the free services do (averages of 113.3 commands per hour versus 86.2 commands per hour) and also run about 1.7 times more questions per hour through their terminals than the free services do (averages of 5.4 questions per hour for the charging services versus 3.2 questions per hour for the free services). This supports the notion that the cost-conscious installations are more productive searchers; however, this data is clouded by the fact that all of the charging installations have high speed terminal equipment, consequently we do not know what contribution to terminal productivity is made by these two separate factors.

6. User Versus Intermediary Searching, and Extent of User Involvement

At the 1974 ASIS annual meeting, Dave McCarn gave a paper which described some of the experiences with MEDLINE searching. In that paper he noted that 75% of the MEDLINE searches were run without the user being present, even though it was his experience that on-line searches took slightly less time to perform when the user was present during the search operation and participated in the search process. This result is contrary to the experience of some other searchers. No data was collected during this project to test this hypothesis, however it is mentioned as another possible factor that might influence terminal productivity.

7. Availability and Use of Analyst Support Tools

Search efficiencies could be influenced by the extent to which analyst support tools (e.g. thesauri, term frequency lists, operating manuals, other authority lists) were available and used by the searchers. We do know that most of the installations had the more important tools, but we do not know the extent to which they were used.

TABLE 17

TERMINAL PRODUCTIVITY RELATED TO COST-CONSCIOUS ATTITUDES OF EACH TERMINAL INSTALLATION

Terminal	Total Number of Commands Entered	Total Number of Hours of Terminal Operation	Total Number of Questions	Average Number of Commands Entered Per Terminal Hour	Average Number of Questions Per Hour
COST RECOVERY OR FULL CHARGING INSTALLATIONS					
4	3,492	21.79	239	160.3	11.1
5	3,470	39.43	112	88.0	2.8
7	6,623	58.68	302	112.9	5.2
All Charging Installations	13,585	119.90	653	113.3	5.4
FREE SERVICE INSTALLATIONS					
2	2,503	28.00	58	89.4	2.1
3	666	7.54	31	88.3	4.1
88	588	8.55	32	68.8	3.7
114	3,045	31.85	113	95.6	3.6
125	1,016	14.79	54	68.7	3.7
All Free Installations	7,818	90.73	288	86.2	3.2
MIXED POLICY (Free to local users, charges for outsiders)					
9	3,635	28.38	70	128.1	2.5
ALL TERMINALS					
	25,038	239.01	1,911	104.8	4.23

G. REFERENCES

1. Summit, Roger. *Remote Information Retrieval Facility*. Palo Alto, Calif. Lockheed Missiles and Space Co., April 1969. 44 pp. NASA CR-1318.

2. Williams, Martha. "Experiences of IIT Research Institute in Operating a Computerized Retrieval System for Searching a Variety of Data Bases," *Information Storage and Retrieval* 8:2 (April 1972) 57-75.

3. Carmon, James. *Annual Report to N.S.F. 1969-70*. Athens, Ga. University of Georgia. Georgia Information Dissemination Center. 1970.

VII. GUIDELINES FOR SEARCHING THE ERIC FILES USING DIALOG

A. INTRODUCTION

In attempting to develop some guidelines for on-line searching of the ERIC data base with DIALOG, consideration was given to the following areas:

-- Pre-Searching Activity (general considerations, procedures, decisions)

-- Terminal Activity (recommended keyboard procedures)

-- Search Strategy (number of terms needed to adequately express each facet or concept of a multi-facet search; methods of limiting quantity of output: effect on relevance).

Each of these areas is treated below as a separate section. Topics in the last two sections were suggested by an informal paper by Charles Missar of the National Institute of Education,[1] and in these sections we look into the quantitative aspects of attempts to increase recall, on the one hand, and to limit quantity of output, on the other hand.

These guidelines are written to incorporate information from many sources, including the findings of this project, comments and suggestions made by search analysts and terminal operators from many search facilities, and comments made at recent ERIC users meetings.

This chapter is not intended to serve as an introduction to DIALOG, or to the ERIC data base. For those topics the reader is referred to Lockheed's Terminal Users Reference Manual,[2] Interchange,[3] and Lockheed's DIALOG Chronolog.[4] This chapter is directed specifically to the use of the ERIC data base as implemented in DIALOG, and is not necessarily generalizable to other data bases or other search systems.

Many of the following recommendations are routine practice for many existing ERIC/DIALOG installations, and even for some other on-line search systems. However, we review them here for completeness and for the benefit of new terminal users.

Our frame of reference is that at present, most DIALOG searchers are acting as intermediaries -- interpreting and acting upon requests received from requestors by mail, by telephone, in person, or through further intermediaries in the field. It is expected that DIALOG searching will continue to be done primarily by trained intermediaries. Some of the points we shall discuss will, however, also be applicable to the work of a requestor searching directly without using an intermediary.

The intermediaries will generally be operating in one of two environments: an information retrieval and dissemination center (where the work of the center is mainly devoted to processing search requests); or a library (where on-line searching is but one of a wide spectrum of reference services provided). At the time of this study most of the installations searching the ERIC files were of the former type. In the future it is quite likely that more libraries will offer on-line searching services as one part of their regular reference services, and more terminals will be installed in offices and departments to provide direct service to end-users.

B. PRE-SEARCH ACTIVITY

Given the above environments, this section will discuss some of the decisions which are made (or should be made) by searchers, consciously or unconsciously, during the pre-search period, i.e. before going to the terminal.

1. Decision: Whether to Go On-Line

The searcher-intermediary should consider whether a particular question could be handled as well manually as on-line. (Note: batch searching is a separate issue that is out of the scope of this study.) The major reason for considering this question is that in some circumstances a manual search may be more cost effective than an on-line search. Furthermore, the requested material may be out of scope of the ERIC file; the moral of this is, "Don't do an on-line search for things that are not in the file."

The environment may affect the decision: a librarian with the printed indexes conveniently at hand might opt for a manual search in some cases; an information center staff member might receive only pre-screened questions which had already passed this decision point; a person with ready access to a terminal but without easy access to the printed indexes might prefer the on-line search in any event; a searcher with no budget restrictions might prefer in all cases to do an on-line search.

To understand the alternatives, consider that there are no multi-year cumulations of the ERIC indexes for some of the search access points. In manually searching the printed ERIC indexes, a searcher must consult annual indexes for each past year of interest, and semiannual or quarterly indexes for the current year, plus the indexes in each issue of the current year not yet cumulated. All of this must be done separately for the RIE and CIJE series of publications. To do a comprehensive single-author search of the printed ERIC indexes as of this report date would require over 10 minutes of manual lookup effort in 25 separate volumes (12 RIE volumes: annual indexes from 1967, plus supplemental issues/25 CIJE volumes: annual indexes from 1969, plus supplemental issues). On the other hand, the DIALOG on-line search provides access to the combined RIE and CIJE files, back to their inception in 1966/7 and 1969, respectively. The RIE and CIJE files are now updated monthly. A single term search for the combined RIE/CIJE file would typically take about two minutes or less of on-line time, especially if a fast search process were used (e.g., BEGIN BYPASS, SELECT term, PRINT).

The only general guideline proposed here is that manual searches should be seriously considered for some types of simple searches, particularly if the installation is very conscious of the costs of on-line service. However, the exact response also depends on what type of simple search is required. For example, for single term searches:

- personal author search. It is probably faster and more cost-effective to search on-line than to manually search through at least 12 separate printed RIE indexes or 25 separate CIJE indexes -- particularly over long time periods.

BEST COPY AVAILABLE

- corporate author search. It may be a toss-up. Because the on-line display of corporate author entries is limited to two 24 character lines, there may be some ambiguity in the displayed items (e.g., as for the several University of California entries shown in Figure 17) that will require more on-line time for citation displays or printouts in order to search the desired institution. In the case of long or complex corporate author names, it might be better to do a manual search. Figure 17 provides some examples of the different forms of entry of identifiers and "publication source" entries, and the effect of truncated index entries (as presented by the EXPAND command), on legibility of corporate author entries. The different forms reflect an area in which the ERIC processing centers have not in the past exerted rigid authority control. The truncation by the EXPAND command is a system feature which is an inconvenience in this regard, and hopefully could be improved upon. The printed source index, Institutional Sources, Statistics and Postings, will provide the accession number as shown in Figure 18 for reports associated with the names of organizations which prepared documents (Institutional Source) or which sponsored the work (Sponsoring Agency) covered in the RIE data base. It is fully cumulated annually and can result in a fast manual search, although yielding only accession numbers. The full text indexing of corporate source entries will provide some on-line advantage here when it becomes available.

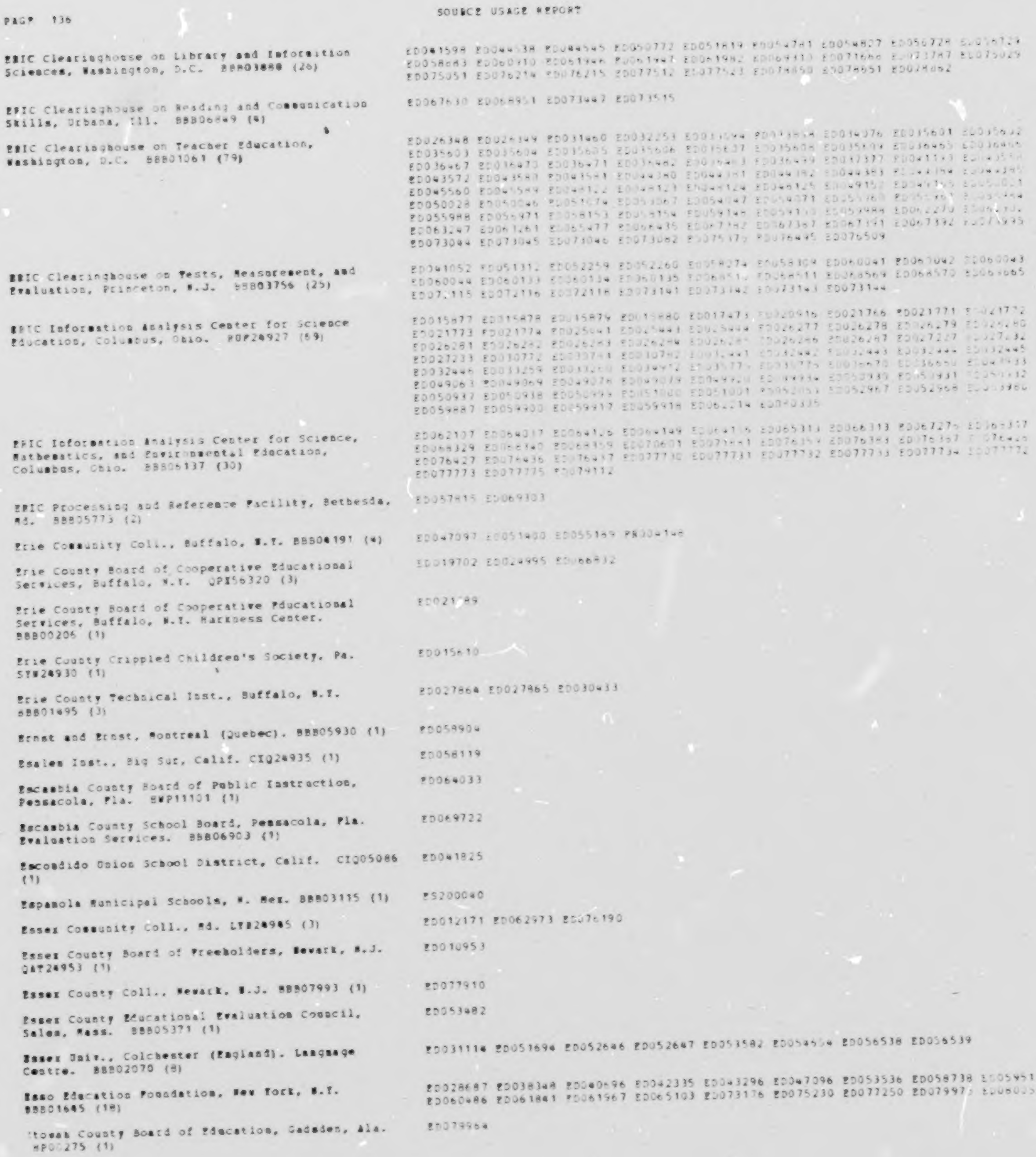
- subject search. In some instances it might be better to do this manually. A cumulative printed index of Descriptors and Identifiers is available as shown in Figure 19. It gives an ED or EJ number for all of the items indexed by that term through April 1973.[5] For some searches, such as these that do not require a search of the most recent material, this might be entirely adequate. However, no abstract or citation is printed by this type of search.

- title search. Title searches can presently only be done manually, using the printed Title Index[6] which is fully cumulated annually and provides title access to the entire RIE report collection through an alphabetic listing of all RIE titles. The DIALOG system presently does not provide a title word search capability for the ERIC data base, however, it is scheduled to be available in January 1975.

- number search. Searches of the RIE data base by report numbers, project numbers, contract numbers, and grant numbers can be done very quickly with the printed ERIC tools, Report/Project Number Index,[7] Contract/Grant Number Index,[8] and Clearinghouse Number to ED Number Cross Reference List.[9] These publications are cumulated through December 1972. All of these files can be searched on-line.

Because some of the single-term searches will yield a large number of retrieved citations, the manual searcher may still be faced with an output task of locating and copying the citations and abstracts from the monthly issues of RIE or CIJE. One alternative to consider here is to search the

BEST COPY AVAILABLE



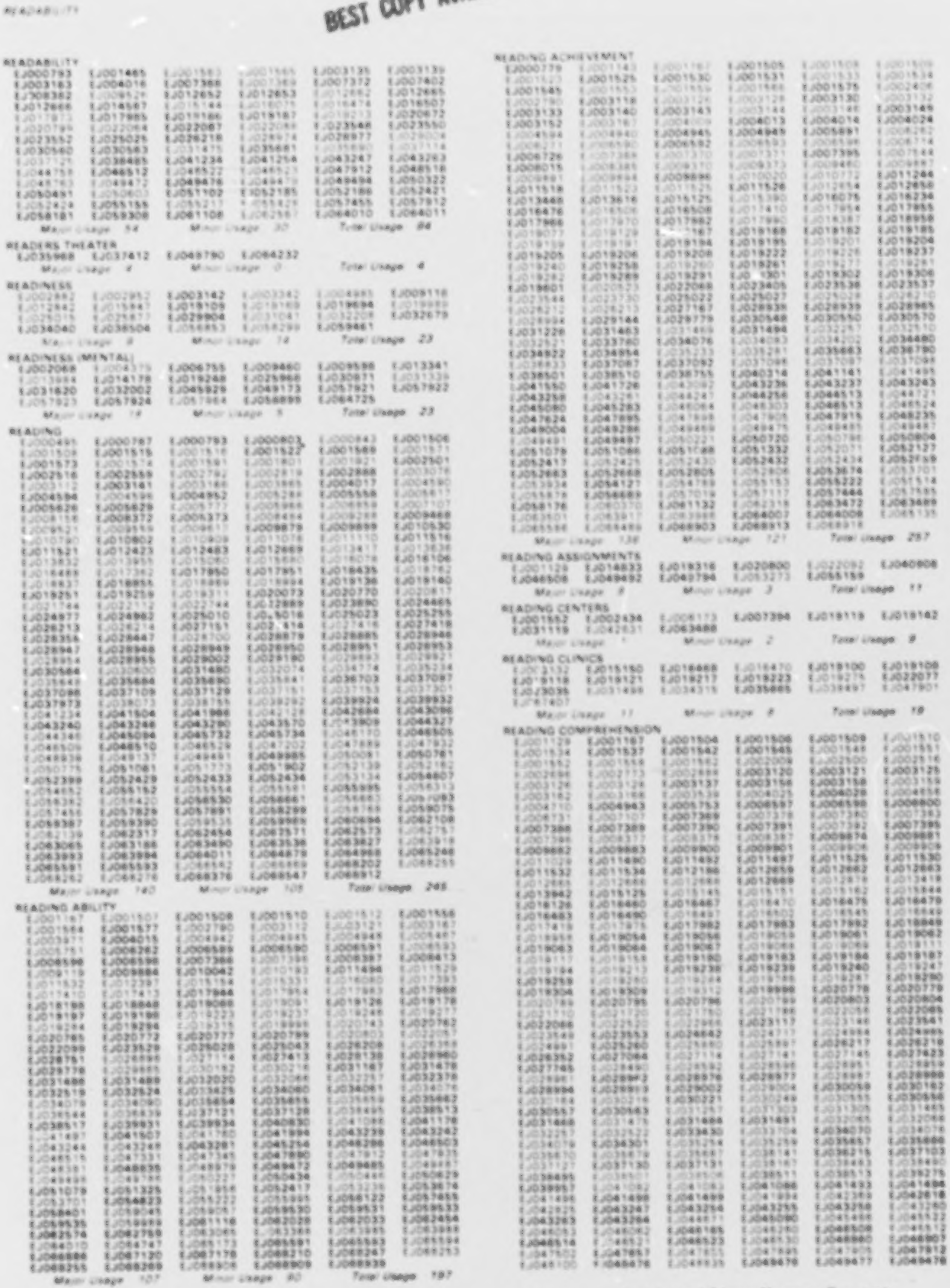
PAGE 136

SOURCE USAGE REPORT

ERIC Clearinghouse on Library and Information Sciences, Washington, D.C. BBB03888 (26)	ED041598 ED044538 ED044545 ED050772 ED051819 ED054781 ED054807 ED056728 ED056729 ED058683 ED060910 ED061946 ED061947 ED061982 ED069313 ED071666 ED073787 ED075029 ED075051 ED076214 ED076215 ED077512 ED077523 ED078650 ED078651 ED078652
ERIC Clearinghouse on Reading and Communication Skills, Urbana, Ill. BBB06849 (4)	ED067630 ED068951 ED073447 ED073515
ERIC Clearinghouse on Teacher Education, Washington, D.C. BBB01061 (79)	ED026348 ED026349 ED031460 ED032253 [illegible] [illegible] ED034976 ED035601 ED035602 ED035603 ED035604 ED035605 ED035606 ED035607 ED035608 ED035609 ED036465 ED036466 ED036467 ED036470 ED036471 ED036482 ED036483 ED036499 ED037377 ED041133 [illegible] ED043572 ED043580 ED043581 ED044380 ED044381 ED044382 ED044383 ED044384 ED044385 ED045560 ED045589 ED048122 ED048123 ED048124 ED048125 ED049152 [illegible] [illegible] ED050028 ED050046 ED051674 ED053067 ED054047 ED054071 ED055760 ED055761 [illegible] ED055988 ED056971 ED058153 ED058154 ED059148 [illegible] ED059988 ED062270 [illegible] ED063247 ED063261 ED065477 ED066435 ED067382 ED067367 ED067391 ED067392 [illegible] ED073044 ED073045 ED073046 ED073082 ED075179 ED076445 ED076509
ERIC Clearinghouse on Tests, Measurement, and Evaluation, Princeton, N.J. BBB03756 (25)	ED041052 ED051312 ED052259 ED052260 ED058274 ED058309 ED060041 ED060042 ED060043 ED060044 ED060133 ED060134 ED060135 ED068510 ED068511 ED068569 ED068570 ED068665 ED072115 ED072116 ED072118 ED073141 ED073142 ED073143 ED073144
ERIC Information Analysis Center for Science Education, Columbus, Ohio. RDF24927 (69)	ED015877 ED015878 ED015879 ED015880 ED017473 ED020916 ED021766 ED021771 ED021772 ED021773 ED021774 ED025441 ED025443 ED025444 ED026277 ED026278 ED026279 ED026280 ED026281 ED026282 ED026283 ED026284 ED026285 ED026286 ED026287 ED027227 ED027232 ED027233 ED030772 [illegible] ED030782 [illegible] ED032442 ED032443 ED032444 ED032445 ED032446 ED033259 ED033260 [illegible] [illegible] [illegible] ED036670 [illegible] [illegible] ED049063 ED049069 ED049078 ED049079 [illegible] [illegible] ED050930 ED050931 ED050932 ED050937 ED050938 [illegible] ED051000 ED051001 [illegible] ED052967 ED052968 [illegible] ED059887 ED059900 ED059917 ED059918 ED062214 ED062335
ERIC Information Analysis Center for Science, Mathematics, and Environmental Education, Columbus, Ohio. BBB06137 (30)	ED062107 ED064017 ED064126 ED064149 [illegible] ED065311 ED066313 ED067275 [illegible] ED068329 ED068340 ED068359 ED070601 [illegible] [illegible] [illegible] [illegible] [illegible] ED076427 ED076436 ED076437 ED077730 ED077731 ED077732 ED077733 ED077734 ED077772 ED077773 ED077775 ED079112
ERIC Processing and Reference Facility, Bethesda, Md. BBB05773 (2)	ED057815 ED069303
Erie Community Coll., Buffalo, N.Y. BBB04191 (4)	ED047097 ED051400 ED055189 PR004148
Erie County Board of Cooperative Educational Services, Buffalo, N.Y. QPX56320 (3)	ED019702 ED024995 ED066832
Erie County Board of Cooperative Educational Services, Buffalo, N.Y. Harkness Center. BBB00206 (1)	ED021789
Erie County Crippled Children's Society, Pa. SYW24930 (1)	ED015610
Erie County Technical Inst., Buffalo, N.Y. BBB01495 (3)	ED027864 ED027865 ED030433
Ernst and Ernst, Montreal (Quebec). BBB05930 (1)	ED058904
Esalen Inst., Big Sur, Calif. CIQ24935 (1)	ED058119
Escambia County Board of Public Instruction, Pensacola, Fla. BWP11101 (1)	ED064033
Escambia County School Board, Pensacola, Fla. Evaluation Services. BBB06903 (1)	ED069722
Escondido Union School District, Calif. CIQ05086 (1)	ED041825
Espanola Municipal Schools, N. Mex. BBB03115 (1)	ES200040
Essex Community Coll., Md. LYB24985 (3)	ED012171 ED062973 ED076190
Essex County Board of Freeholders, Newark, N.J. QAT24953 (1)	ED010953
Essex County Coll., Newark, N.J. BBB07993 (1)	ED077910
Essex County Educational Evaluation Council, Salem, Mass. BBB05371 (1)	ED053482
Essex Univ., Colchester (England). Language Centre. BBB02070 (8)	ED031114 ED051694 ED052646 ED052647 ED053582 ED054654 ED056538 ED056539
Esso Education Foundation, New York, N.Y. BBB01645 (18)	ED028687 ED038348 ED040696 ED042335 ED043296 ED047096 ED053536 ED058738 ED059513 ED060486 ED061841 ED061967 ED065103 ED073176 ED075230 ED077250 ED079973 ED080352
towah County Board of Education, Gadsden, Ala. HPO0275 (1)	ED079964

Fig. 18
Sample Page from Institutional Sources, Statistics and Postings

BEST COPY AVAILABLE



1188 Fig. 19. Sample Page from Macmillan Descriptor and Identifier Usage Report term in the printed tools in order to determine the number of retrieved citations. For more than a very few citations it would clearly be easier and less expensive over all to search and print the citations and abstracts with the computer approach.

On-line searches are most appropriate for multi-term or multi-aspect searches when an intersection of two or more terms or groups of terms is required to answer a question. An intersection is defined here as the combination of two or more terms using Boolean AND logic.

The relative merits of on-line vs. manual searching are less easily seen for questions where a few terms in a simple OR relationship are required. Often such a search could be carried out rather easily, though not as quickly, using conventional printed indexes.

Given the present DIALOG system, ERIC data-base, and printed ERIC indexes, there are several points in favor of doing an on-line search instead of a manual search in the printed ERIC indexes:

-- The search is done in one operation, rather than having to be repeated over many printed index volumes.

-- Both major descriptors (those marked with asterisks on the computer printout copy and in the printed indexes) as well as minor descriptors (unmarked descriptors, which are in the machine file but omitted from the printed indexes), may be searched on-line. This means that in cases where a requester desires to see all citations which have been indexed by a specific term, a computer search would be appropriate; in fact this search could not be done with the printed indexes. A more detailed discussion of the major/minor descriptor values used in ERIC indexing is given in the later section on Methods of Limiting Quantity of Output.

-- Identifiers, which do not appear in the printed indexes, but are contained in the machine file, may be searched. Identifiers are often used when a term is new and has not yet graduated to descriptor status.

-- In cases where the printed indexes are not readily available, on-line searching will probably be more convenient.

-- Title word searching (if and when added) will be an on-line capability with no manual equivalent.

-- Stem searching will be an on-line convenience when searching some terms (e.g., computation, computational, computed, computer, computer-, computerized, computers, computing).

-- After identifying the relevant ED or EJ numbers, a computer-printed bibliography can be obtained faster, more conveniently, and at significantly less cost than the alternative manual process of locating each citation in the appropriate RIE or CIJE monthly volume and then copying the selected citations and abstracts. This output effort can be a significant factor when the typical search results in 50-100 citations.

BEST COPY AVAILABLE

2. Decision: Whether to Use Printed Analyst Support Tools Before Going to the Terminal

Most terminal installations now argue that searchers should do some planning and analysis work on most searches before they go to the terminal. This includes at least the preliminary identification of the major facets to be searched, the logical relationships between these facets, and some initial search terms. Some installations consider it essential to use some sort of form sheet to work up the search specifics prior to searching. During this pre-search activity, the analyst may benefit from one or more of the analyst support tools discussed below.

a. The ERIC Thesaurus

The ERIC system performs subject indexing of incoming items in conjunction with a controlled vocabulary that was developed at the beginning of the ERIC system, and has been carefully and closely maintained since then. This indexing vocabulary of over 7,500 terms is published as the ERIC Thesaurus,[10] and re-issued in an updated form annually. A sample page from that Thesaurus is shown in Figure 20. In the ERIC system, all of the subject index terms listed in the Thesaurus are defined to be Descriptors, and that terminology and distinction is used in this report. Other uncontrolled subject index terms may also be assigned to each incoming item, particularly for specific names (e.g., Bronx Zoo, B6700, Captain Kangaroo) or terms that are not likely to result in enough postings to make it worthwhile to include in the Thesaurus (e.g., caper, cardiac, cats). In the ERIC system, these terms, over 22,000 of them, are defined to be Identifiers. It is possible that the same term might be used in some earlier items as an Identifier, and in a later item as a Descriptor. There is an average of 10.46 Descriptors per RIE accession, and 6.88 Descriptors per CIJE accession. There is an average of 1.75 Identifiers per RIE accession, and 1.37 Identifiers per CIJE accession.

Searchers should consider whether, and how much, they should use the ERIC Thesaurus before going to the terminal, since the thesaurus is also available for on-line display and may be used efficiently there. The searcher may choose between the following alternatives:

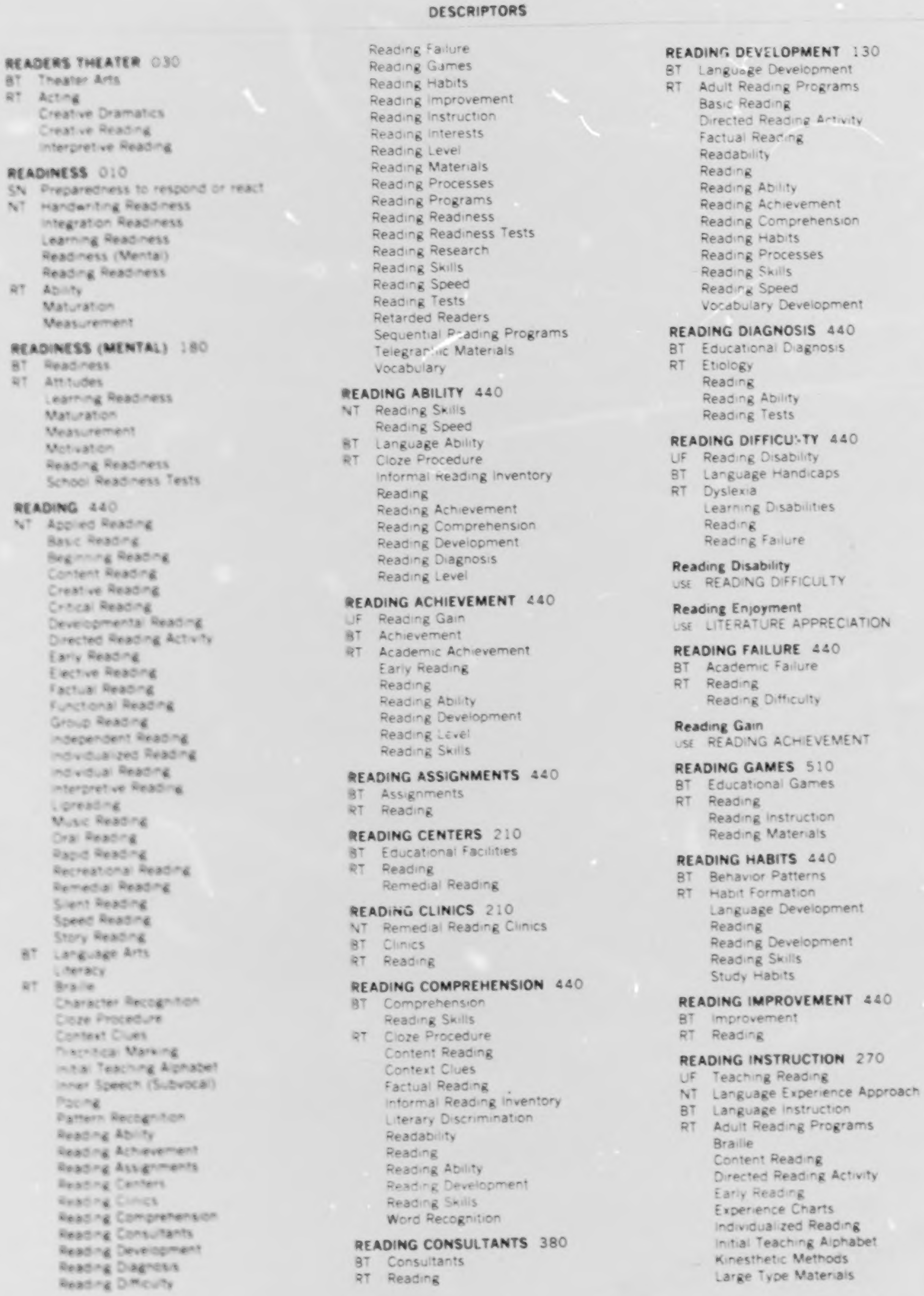
1) Using the printed Thesaurus before going to the terminal, and not using the Thesaurus on-line. (This may be cumbersome.)

2) Using the on-line Thesaurus with no use of the printed Thesaurus. (This may suffice for experienced searchers.)

3) Using the printed Thesaurus to sketch out the proposed search, and then the on-line Thesaurus for convenient selection of terms. (This may work well for less experienced searchers.)

4) Not using the Thesaurus at all. (This is not advisable.)

BEST COPY AVAILABLE



DESCRIPTORS

READERS THEATER 030
BT Theater Arts
RT Acting
Creative Dramatics
Creative Reading
Interpretive Reading

READINESS 010
SN Preparedness to respond or react
NT Handwriting Readiness
Integration Readiness
Learning Readiness
Readiness (Mental)
Reading Readiness
RT Ability
Maturation
Measurement

READINESS (MENTAL) 180
BT Readiness
RT Attitudes
Learning Readiness
Maturation
Measurement
Motivation
Reading Readiness
School Readiness Tests

READING 440
NT Applied Reading
Basic Reading
Beginning Reading
Content Reading
Creative Reading
Critical Reading
Developmental Reading
Directed Reading Activity
Early Reading
Elective Reading
Factual Reading
Functional Reading
Group Reading
Independent Reading
Individualized Reading
Individual Reading
Interpretive Reading
Lipreading
Music Reading
Oral Reading
Rapid Reading
Recreational Reading
Remedial Reading
Silent Reading
Speed Reading
Story Reading
BT Language Arts
Literacy
RT Braille
Character Recognition
Cloze Procedure
Context Clues
Diacritical Marking
Initial Teaching Alphabet
Inner Speech (Subvocal)
Pacing
Pattern Recognition
Reading Ability
Reading Achievement
Reading Assignments
Reading Centers
Reading Clinics
Reading Comprehension
Reading Consultants
Reading Development
Reading Diagnosis
Reading Difficulty
Reading Failure
Reading Games
Reading Habits
Reading Improvement
Reading Instruction
Reading Interests
Reading Level
Reading Materials
Reading Processes
Reading Programs
Reading Readiness
Reading Readiness Tests
Reading Research
Reading Skills
Reading Speed
Reading Tests
Retarded Readers
Sequential Reading Programs
Telegraphic Materials
Vocabulary

READING ABILITY 440
NT Reading Skills
Reading Speed
BT Language Ability
RT Cloze Procedure
Informal Reading Inventory
Reading
Reading Achievement
Reading Comprehension
Reading Development
Reading Diagnosis
Reading Level

READING ACHIEVEMENT 440
UF Reading Gain
BT Achievement
RT Academic Achievement
Early Reading
Reading
Reading Ability
Reading Development
Reading Level
Reading Skills

READING ASSIGNMENTS 440
BT Assignments
RT Reading

READING CENTERS 210
BT Educational Facilities
RT Reading
Remedial Reading

READING CLINICS 210
NT Remedial Reading Clinics
BT Clinics
RT Reading

READING COMPREHENSION 440
BT Comprehension
Reading Skills
RT Cloze Procedure
Content Reading
Context Clues
Factual Reading
Informal Reading Inventory
Literary Discrimination
Readability
Reading
Reading Ability
Reading Development
Reading Skills
Word Recognition

READING CONSULTANTS 380
BT Consultants
RT Reading

READING DEVELOPMENT 130
BT Language Development
RT Adult Reading Programs
Basic Reading
Directed Reading Activity
Factual Reading
Readability
Reading
Reading Ability
Reading Achievement
Reading Comprehension
Reading Habits
Reading Processes
Reading Skills
Reading Speed
Vocabulary Development

READING DIAGNOSIS 440
BT Educational Diagnosis
RT Etiology
Reading
Reading Ability
Reading Tests

READING DIFFICULTY 440
UF Reading Disability
BT Language Handicaps
RT Dyslexia
Learning Disabilities
Reading
Reading Failure

Reading Disability
USE READING DIFFICULTY

Reading Enjoyment
USE LITERATURE APPRECIATION

READING FAILURE 440
BT Academic Failure
RT Reading
Reading Difficulty

Reading Gain
USE READING ACHIEVEMENT

READING GAMES 510
BT Educational Games
RT Reading
Reading Instruction
Reading Materials

READING HABITS 440
BT Behavior Patterns
RT Habit Formation
Language Development
Reading
Reading Development
Reading Skills
Study Habits

READING IMPROVEMENT 440
BT Improvement
RT Reading

READING INSTRUCTION 270
UF Teaching Reading
NT Language Experience Approach
BT Language Instruction
RT Adult Reading Programs
Braille
Content Reading
Directed Reading Activity
Early Reading
Experience Charts
Individualized Reading
Initial Teaching Alphabet
Kinesthetic Methods
Large Type Materials

Fig. 20. Sample Page from ERIC Thesaurus

Unless the searchers are very experienced and familiar with the subject matter of the particular question at hand, some initial use of the printed thesaurus is advisable. The search should be sketched out in advance of terminal use, showing the facets which are to be developed, and delineating the logical relationship between facets. (By facet we mean a term or group of terms which expresses one aspect of a search topic. Typical ERIC search facets would be age/grade level (e.g., high school, secondary school), subject field (e.g., science, mathematics), and approach (e.g., audiovisual instruction)). Each facet can usually be expressed by several roughly equivalent terms which are ORed together; two or three facets are typically ANDed together, forming an intersected set.) In our opinion, no more than a sketch is needed at this point; it would be cumbersome to write out great lists of terms by hand. But the skeleton set of terms provides the searcher at the terminal with:

-- Starting points for use of the EXPAND command;

-- Memory jogs in case some desired terms do not show up in the Thesaurus as "related terms" during the course of the search;

BEST COPY AVAILABLE

-- Advance planning time for handling terms which need special treatment -- e.g., terms which themselves include more than one facet of the planned search ("secondary school science" as opposed to "science instruction" and "secondary schools.." etc.).

-- Information regarding which terms should perhaps not be keyed on-line for a descriptor search (because the term is absent from the Thesaurus). Thesaurus "Use For" terms are a good example of this. The Use For relationship indicates that one term should be used for indexing or searching instead of another. The Use For terms are given in the printed Thesaurus but are not given in the on-line display.

b. Term Frequency Lists

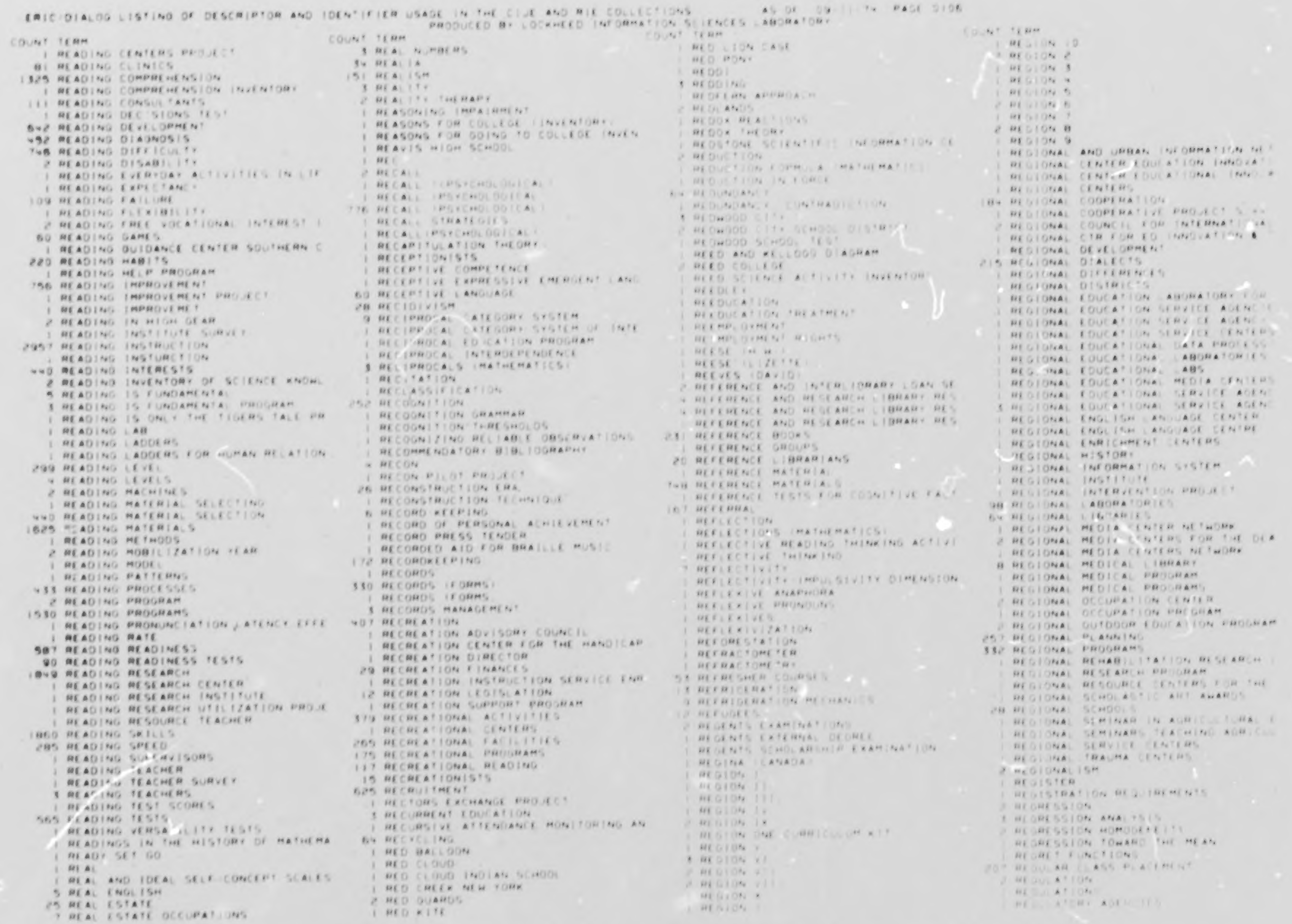
A helpful tool for identifying situations in which there may be many postings for a given term, is a cumulative term frequency list which shows how many file items have been indexed by each term used in the ERIC system. Such a listing can also identify terms that might have been considered for searching, but should not be keyed in on-line because no items (as of the date of the list) are indexed by that term. Term frequency lists for the ERIC data base have been prepared by several organizations, and can generally be obtained at very little cost from the originating organization. A brief description of several of these lists is given in Table 18. Sample pages from two of these lists are given in Figures 19 and 21.

The Lockheed list[11] gives frequencies of both Descriptors and Identifiers, merged together in one alphabetical listing just as they are displayed on-line by the EXPAND command. The Macmillan report[5] gives term frequencies as well as ED or EJ numbers of items indexed by those terms, but divides the report into Descriptor and Identifier sections. The North Carolina report gives term frequencies only.[12]

TABLE 18

TERM FREQUENCY LISTS FOR THE ERIC DATA BASE

Source	Date of Cumulation	DATA INCLUDED			
		Descriptors	Identifiers	Posting Frequency	Accession Number
Lockheed	Aug., 1974	X (merged list)	X	X	
Macmillan Information	Apr., 1973	X (separate lists for Descriptors and Identifiers, divided by source-- RIE or CIJE)	X	X	X
North Carolina Board of Science and Technology	Sept., 1974	X (merged list of Descriptors and Identifiers, divided by source-- RIE or CIJE)	X	X	



```
ERIC/DIALOG LISTING OF DESCRIPTOR AND IDENTIFIER USAGE IN THE CIJE AND RIE COLLECTIONS        AS OF  09/11/74   PAGE 0108
                                    PRODUCED BY LOCKHEED INFORMATION SCIENCES LABORATORY
COUNT TERM
    1 READING CENTERS PROJECT
   81 READING CLINICS
 1329 READING COMPREHENSION
    1 READING COMPREHENSION INVENTORY
  111 READING CONSULTANTS
    1 READING DECISIONS TEST
  842 READING DEVELOPMENT
  492 READING DIAGNOSIS
  748 READING DIFFICULTY
    2 READING DISABILITY
    1 READING EVERYDAY ACTIVITIES IN LIF
    1 READING EXPECTANCY
  109 READING FAILURE
    1 READING FLEXIBILITY
    2 READING FREE VOCATIONAL INTEREST I
   60 READING GAMES
    1 READING GUIDANCE CENTER SOUTHERN C
  220 READING HABITS
    1 READING HELP PROGRAM
  756 READING IMPROVEMENT
    1 READING IMPROVEMENT PROJECT
    1 READING IMPROVEMET
    2 READING IN HIGH GEAR
    1 READING INSTITUTE SURVEY
 2957 READING INSTRUCTION
    1 READING INSTURCTION
  440 READING INTERESTS
    2 READING INVENTORY OF SCIENCE KNOWL
    5 READING IS FUNDAMENTAL
    3 READING IS FUNDAMENTAL PROGRAM
    1 READING IS ONLY THE TIGERS TALE PR
    1 READING LAB
    1 READING LADDERS
    1 READING LADDERS FOR HUMAN RELATION
  299 READING LEVEL
    4 READING LEVELS
    2 READING MACHINES
    1 READING MATERIAL SELECTING
  440 READING MATERIAL SELECTION
 1625 READING MATERIALS
    1 READING METHODS
    2 READING MOBILIZATION YEAR
    1 READING MODEL
    1 READING PATTERNS
  433 READING PROCESSES
    2 READING PROGRAM
 1530 READING PROGRAMS
    1 READING PRONUNCIATION LATENCY EFFE
    1 READING RATE
  587 READING READINESS
   90 READING READINESS TESTS
 1849 READING RESEARCH
    1 READING RESEARCH CENTER
    1 READING RESEARCH INSTITUTE
    1 READING RESEARCH UTILIZATION PROJE
    1 READING RESOURCE TEACHER
 1860 READING SKILLS
  285 READING SPEED
    1 READING SUPERVISORS
    1 READING TEACHER
    1 READING TEACHER SURVEY
    3 READING TEACHERS
    1 READING TEST SCORES
  565 READING TESTS
    1 READING VERSATILITY TESTS
    1 READINGS IN THE HISTORY OF MATHEMA
    1 READY SET GO
    1 REAL
    1 REAL AND IDEAL SELF CONCEPT SCALES
    5 REAL ENGLISH
   25 REAL ESTATE
    7 REAL ESTATE OCCUPATIONS

COUNT TERM
    3 REAL NUMBERS
   54 REALIA
  151 REALISM
    3 REALITY
    2 REALITY THERAPY
    1 REASONING IMPAIRMENT
    1 REASONS FOR COLLEGE (INVENTORY)
    1 REASONS FOR GOING TO COLLEGE INVEN
    1 REAVIS HIGH SCHOOL
    1 REC
    2 RECALL
    1 RECALL ((PSYCHOLOGICAL)
    1 RECALL (PSYCHOLOGICAL
  776 RECALL (PSYCHOLOGICAL)
    1 RECALL STRATEGIES
    1 RECALL(PSYCHOLOGICAL)
    1 RECAPITULATION THEORY
    1 RECEPTIONISTS
    1 RECEPTIVE COMPETENCE
    1 RECEPTIVE EXPRESSIVE EMERGENT LANG
   60 RECEPTIVE LANGUAGE
   28 RECIDIVISM
    9 RECIPROCAL CATEGORY SYSTEM
    1 RECIPROCAL CATEGORY SYSTEM OF INTE
    1 RECIPROCAL EDUCATION PROGRAM
    1 RECIPROCAL INTERDEPENDENCE
    3 RECIPROCALS (MATHEMATICS)
    1 RECITATION
    1 RECLASSIFICATION
  252 RECOGNITION
    1 RECOGNITION GRAMMAR
    1 RECOGNITION THRESHOLDS
    1 RECOGNIZING RELIABLE OBSERVATIONS
    1 RECOMMENDATORY BIBLIOGRAPHY
    4 RECON
    1 RECON PILOT PROJECT
   26 RECONSTRUCTION ERA
    1 RECONSTRUCTION TECHNIQUE
    6 RECORD KEEPING
    1 RECORD OF PERSONAL ACHIEVEMENT
    1 RECORD PRESS TENDER
    1 RECORDED AID FOR BRAILLE MUSIC
  172 RECORDKEEPING
    1 RECORDS
  330 RECORDS (FORMS)
    1 RECORDS (FORMS
    3 RECORDS MANAGEMENT
  407 RECREATION
    1 RECREATION ADVISORY COUNCIL
    1 RECREATION CENTER FOR THE HANDICAP
    1 RECREATION DIRECTOR
   29 RECREATION FINANCES
    1 RECREATION INSTRUCTION SERVICE ENR
   12 RECREATION LEGISLATION
    1 RECREATION SUPPORT PROGRAM
  379 RECREATIONAL ACTIVITIES
    1 RECREATIONAL CENTERS
  265 RECREATIONAL FACILITIES
  175 RECREATIONAL PROGRAMS
  117 RECREATIONAL READING
   15 RECREATIONISTS
  625 RECRUITMENT
    1 RECTORS EXCHANGE PROJECT
    3 RECURRENT EDUCATION
    1 RECURSIVE ATTENDANCE MONITORING AN
   64 RECYCLING
    1 RED BALLOON
    1 RED CLOUD
    1 RED CLOUD INDIAN SCHOOL
    1 RED CREEK NEW YORK
    2 RED GUARDS
    1 RED KITE

COUNT TERM
    1 RED LION CASE
    1 RED PONY
    1 REDDI
    3 REDDING
    1 REDFERN APPROACH
    2 REDLANDS
    1 REDOX REACTIONS
    1 REDOX THEORY
    1 REDSTONE SCIENTIFIC INFORMATION CE
    2 REDUCTION
    1 REDUCTION FORMULA (MATHEMATICS)
    1 REDUCTION IN FORCE
   64 REDUNDANCY
    1 REDUNDANCY  CONTRADICTION
    3 REDWOOD CITY
    2 REDWOOD CITY SCHOOL DISTRICT
    1 REDWOOD SCHOOL TEST
    1 REED AND KELLOGG DIAGRAM
    2 REED COLLEGE
    1 REED SCIENCE ACTIVITY INVENTORY
    1 REEDLEY
    1 REEDUCATION
    1 REEDUCATION TREATMENT
    1 REEMPLOYMENT
    1 REEMPLOYMENT RIGHTS
    1 REESE (H W)
    1 REESE (LIZETTE)
    1 REEVES (DAVID)
    2 REFERENCE AND INTERLIBRARY LOAN SE
    4 REFERENCE AND RESEARCH LIBRARY RES
    4 REFERENCE AND RESEARCH LIBRARY RES
    1 REFERENCE AND RESEARCH LIBRARY RES
  231 REFERENCE BOOKS
    1 REFERENCE GROUPS
   20 REFERENCE LIBRARIANS
    1 REFERENCE MATERIAL
  748 REFERENCE MATERIALS
    1 REFERENCE TESTS FOR COGNITIVE FACT
  167 REFERRAL
    1 REFLECTION
    1 REFLECTIONS (MATHEMATICS)
    1 REFLECTIVE READING THINKING ACTIVI
    1 REFLECTIVE THINKING
    7 REFLECTIVITY
    1 REFLECTIVITY IMPULSIVITY DIMENSION
    1 REFLEXIVE ANAPHORA
    1 REFLEXIVE PRONOUNS
    1 REFLEXIVES
    1 REFLEXIVIZATION
    1 REFORESTATION
    1 REFRACTOMETER
    1 REFRACTOMETRY
   53 REFRESHER COURSES
   13 REFRIGERATION
    9 REFRIGERATION MECHANICS
   12 REFUGEES
    2 REGENTS EXAMINATIONS
    1 REGENTS EXTERNAL DEGREE
    1 REGENTS SCHOLARSHIP EXAMINATION
    1 REGINA (CANADA)
    1 REGION I
    1 REGION II
    1 REGION III
    1 REGION IV
    2 REGION IX
    1 REGION ONE CURRICULUM KIT
    1 REGION V
    4 REGION VI
    2 REGION VII
    2 REGION VIII
    1 REGION X
    1 REGION 1

COUNT TERM
    1 REGION 10
    3 REGION 2
    1 REGION 3
    1 REGION 4
    1 REGION 5
    2 REGION 6
    1 REGION 7
    2 REGION 8
    1 REGION 9
    1 REGIONAL AND URBAN INFORMATION NET
    1 REGIONAL CENTER EDUCATION INNOVATI
    1 REGIONAL CENTER EDUCATIONAL INNOVA
    1 REGIONAL CENTERS
  184 REGIONAL COOPERATION
    1 REGIONAL COOPERATIVE PROJECT S KA
    2 REGIONAL COUNCIL FOR INTERNATIONAL
    1 REGIONAL CTR FOR ED INNOVATION 6
    1 REGIONAL DEVELOPMENT
  215 REGIONAL DIALECTS
    1 REGIONAL DIFFERENCES
    1 REGIONAL DISTRICTS
    1 REGIONAL EDUCATION LABORATORY FOR
    1 REGIONAL EDUCATION SERVICE AGENCIE
    1 REGIONAL EDUCATION SERV CE AGENC
    1 REGIONAL EDUCATION SERVICE CENTERS
    1 REGIONAL EDUCATIONAL DATA PROCESS
    1 REGIONAL EDUCATIONAL LABORATORIES
    1 REGIONAL EDUCATIONAL LABS
    1 REGIONAL EDUCATIONAL MEDIA CENTERS
    1 REGIONAL EDUCATIONAL SERVICE AGEN
    3 REGIONAL EDUCATIONAL SERVICE AGEN
    1 REGIONAL ENGLISH LANGUAGE CENTER
    1 REGIONAL ENGLISH LANGUAGE CENTRE
    1 REGIONAL ENRICHMENT CENTERS
    1 REGIONAL HISTORY
    1 REGIONAL INFORMATION SYSTEM
    1 REGIONAL INSTITUTE
    1 REGIONAL INTERVENTION PROJECT
   98 REGIONAL LABORATORIES
   64 REGIONAL LIBRARIES
    1 REGIONAL MEDIA CENTER NETWORK
    2 REGIONAL MEDIA CENTERS FOR THE DEA
    1 REGIONAL MEDIA CENTERS NETWORK
    8 REGIONAL MEDICAL LIBRARY
    1 REGIONAL MEDICAL PROGRAM
    1 REGIONAL MEDICAL PROGRAMS
    2 REGIONAL OCCUPATION CENTER
    1 REGIONAL OCCUPATION PROGRAM
    2 REGIONAL OUTDOOR EDUCATION PROGRAM
  257 REGIONAL PLANNING
  332 REGIONAL PROGRAMS
    1 REGIONAL REHABILITATION RESEARCH I
    1 REGIONAL RESEARCH PROGRAM
    1 REGIONAL RESOURCE CENTERS FOR THE
    1 REGIONAL SCHOLASTIC ART AWARDS
   28 REGIONAL SCHOOLS
    1 REGIONAL SEMINAR IN AGRICULTURAL E
    1 REGIONAL SEMINARS TEACHING AGRICU
    1 REGIONAL SERVICE CENTERS
    1 REGIONAL TRAUMA CENTERS
    2 REGIONALISM
    1 REGISTER
    1 REGISTRATION REQUIREMENTS
    2 REGRESSION
    3 REGRESSION ANALYSIS
    2 REGRESSION HOMOGENEITY
    1 REGRESSION TOWARD THE MEAN
    1 REGRET FUNCTIONS
  207 REGULAR CLASS PLACEMENT
    2 REGULATION
    1 REGULATIONS
    1 REGULATORY AGENCIES
```

Fig. 21. Sample Page from Lockheed Term Frequency List

BEST COPY AVAILABLE

c. Other Printed Aids

Another useful printed publication which should be mentioned is the ERIC Processing Manual,[13] which contains information about ERIC indexing and other characteristics of the data base.

3. Decision: Whether to Use SEARCH SAVE

The SEARCH SAVE feature, which is provided by the DIALOG system for the ERIC files, enables searchers to store search statements for later execution with the same or another search request. This feature provides an easy and time-saving way to handle commonly-recurring search facets (e.g., a school grade level) instead of reconstructing them each time they are needed. Some facets such as elementary/secondary education might require 35 or more terms for a complete description; it would be a terrible inconvenience to have to re-key those terms every time that facet was used in a search.

An example of such a saved facet is shown in Figure 22. The SEARCH SAVE file may be thought of as analogous to a collection of computer sub-programs which can be called up by a programmer when needed. A given installation may wish to create its own library of SEARCH SAVES for its own subject areas or repeating concepts. An installation may also use already existing ones, by consulting the list of SEARCH SAVES published by Lockheed,[14] and illustrated earlier in Figure 22. As a side comment here, the usefulness of the SEARCH SAVE list would be enhanced by a title index, and perhaps a keyword index. The SEARCH SAVE feature is intended to be used as a basis for current awareness searching (i.e., SDI) for a given profile. At the time of this report such current awareness searching was implemented only on the Predicasts data base, and no date had been announced yet regarding its use with the ERIC data base.

Before going to the terminal, the searcher should note the file number of any SEARCH SAVE to be used. If no existing SEARCH SAVE is exactly right, but one is needed, the searcher should plan to create the SEARCH SAVE as a separate step.

A SEARCH SAVE is stored when the command END/SAVE, or =/SAVE is issued. At this time the DIALOG system responds, on the terminal, with a 2-character number, such as 6G. The searcher must record this number, by keeping the terminal's printout or by writing the number down, in order to be able to later RECALL the search. Unfortunately the number is not printed on the search history which accompanies any off-line printed citations. It would be helpful if Lockheed would incorporate the number of any newly-created SEARCH SAVE into the off-line printed search history.

A saved search may be recalled and used by giving a .RECALL nn command, where nn is the previously issued SEARCH SAVE number, followed by the command .EXECUTE (n). The descriptor postings are newly derived for the sets specified. The saved search executes to the end, or executes the set number specified, with all its previously defined component sets.

```
USER SER#   DATE     NAME               TITLE

   7    N 12/13/72 CLAY                SEARCH SAVE/HIGH SCHOOLS
     SET COMMAND

          SEARCH SAVE/HIGH SCHOOLS
          CLAY
          SMERC
          SAN MATEO
          1
       1 #HIGH SCHOOLS
       2 #SENIOR HIGH SCHOOLS
       3 #SECONDARY GRADES
       4 #SECONDARY SCHOOLS
       5 #SECONDARY SCHOOL STUDENTS
       6 #GRADE 9
       7 #GRADE 10
       8 #GRADE 11
       9 #GRADE 12
      10 #HIGH SCHOOL STUDENTS
      11 #HIGH SCHOOL CURRICULUM
      12 #HIGH SCHOOL ROLE
      13 #SECONDARY EDUCATION
      14 $1-13/+
         =/SAVE
```

BEST COPY AVAILABLE

Fig. 22a. Example of a Saved Search

```
SET ITEMS DESCRIPTION

25 11863 SERIAL NO.: N/
```

Fig. 22b. Example of Message Reporting Execution of Search Save N, Shown Above

When a saved search is executed, only its number is reported on the search history, as shown in Figure 22. A much more intelligible search history would result if the title of the saved and recalled search were given as well.

The SEARCH SAVE feature was announced in Fall 1972. As shown by the RECALL and EXECUTE command use data in Table 6, it does not appear to have been used very extensively by many terminal installations, and may not have contributed significantly to the performance of the installations during the 15-day period that we examined closely.

4. Advance Determination of Possible Ways to Limit Output

It should be ascertained in advance (while discussing the search topic with the requestor, if possible), whether a broad or narrow search is desired by the requestor; how many citations are desired (or expected); and whether a limitation by date or other criteria would be acceptable if too many citations are retrieved. How many citations are "too many" varies with the individual; most installations have a working assumption that a number of citations from 50 to 100 is appropriate, and more than this number is too many. A few installations feel that most of their users do not need or want more than 5-10 citations.

Many different criteria can be used with the ERIC file as a basis for limiting the output on something other than a subject basis. Examples of limits that can be used are:

- date (of publication, of ERIC accession)
- contributing ERIC clearinghouse (e.g., EC, IR)
- ED versus EJ publication (ED only, EJ only)
- type of publication (state-of-the-art review, annotated bibliography)
- availability of the cited publication through the ERIC Document Reproduction System
- total number of citations to be printed.

These parameters are often built into search request forms. Sometimes additional limiting facets can be specified, e.g., "curriculum work only", "evaluations only". A written statement of the search request should be obtained whenever possible; such written statements often provide clues which can be helpful if the search does not proceed as expected at the terminal.

BEST COPY AVAILABLE

C. TERMINAL ACTIVITY

1. Equipment Considerations

ERIC/DIALOG may be accessed using a number of different equipment configurations (e.g., high-speed or low-speed terminals, high speed dedicated phone lines or lower speed dial-up phone lines; CRT (cathode ray tube) or hard-copy terminals or combinations of these). These considerations were discussed in more detail in an earlier section of this report. During this study we used primarily a high speed (480 characters/second) leased line and CRT terminal, with an auxiliary hard-copy printer; we also used a slow speed dial-up hard-copy terminal. Both types of configuration performed satisfactorily for us.

In considering whether or not to use a CRT-only terminal, a hard-copy-only terminal, or a CRT terminal supplemented by hard-copy printout of selected pieces of information, the following points should be considered.

Hard copy output of the terminal has several advantages:

-- Useful in tracing and recording previous steps in search execution (This may be done on a CRT by the DISPLAY SEARCH HISTORY command.)

-- Can be used for direct printing of retrieved citations at the terminal

-- Provides a printed record of file numbers of saved searches (see previous section on using the SEARCH SAVE)

-- Provides an immediate printed record of the elapsed search time that can be used for charging and cost accounting purposes for those installations that recover costs by service charges.

On a configuration that has both a high-speed CRT and an auxiliary printer, the printer is usually used only to print DIALOG commands and desired citations, thus reducing the volume of terminal printing activity. On hard-copy-only terminals, all DIALOG responses are printed out; this may consume a considerable amount of paper (and make a considerable amount of noise if mechanical printers are used), especially if the EXPAND command is used.

Disadvantages of hard-copy-only terminals are:

-- Uses a lot of paper, especially if EXPAND commands are used; hence the use of this command might tend to be discouraged

-- Slower speed than CRT terminals (in characters per minute) for most types of hard copy terminals

-- May be noisier than the other alternatives

-- If a high speed terminal is desired (e.g., 480 characters /sec.), it is generally more expensive (for both equipment and supplies) to use high speed printing equipment than CRT equipment.

Because Lockheed's response time in delivering off-line printed citations is so fast (citations are printed off-line in the early shift of the morning following Lockheed's receipt of a PRINT command, and sometimes the same day and then sent Air Mail), the time advantage gained by printing citations at the terminal is slight (only a few days); however, having a paper copy can be useful for recordkeeping and reference purposes.

We favor a CRT terminal installation that includes a hard-copy feature, but realize that this may not be cost effective for some other installations. The search speed can be improved through the use of high-speed communications and display equipment, and this should be considered for installations which do a large volume of searching. Test data reported in earlier sections of this report showed that a significantly larger volume of work (questions or searches per hour, commands per hour) was passed through the high speed terminals for the same unit of time. The high speed terminals can be cost effective at moderate volumes of search activity and provide considerable cost savings at high volumes and can be justified from only their fixed communication cost.

2. Recommended Keyboard Procedures

In the previous section we discussed some aspects of search negotiation and preparation which may take place before the searcher goes to the terminal. In this section we will discuss the procedures which may be followed by the searcher at the terminal. It is assumed that the installation will have good sign on and sign off procedures to avoid the charges for terminal time while the searcher takes a break, is interrupted for any significant period of time, or walks away from the terminal and forgets to disconnect the terminal from the system.

In most installations under study here there are probably some tacit assumptions about the frame of reference of the searches being done. One assumption which has a direct effect on activity at the terminal relates to requestor involvement: Is the requestor assumed to be interested in, and capable of understanding the search logic which produces the list of citations? Or is the requestor assumed to be interested only in the output, and not at all in the process? Is the search process iterative with respect to a given request? Or is iteration limited to processing new search topics for a given requestor? The project team assumed that most searches are "one-shot" efforts, not expected to be revised or re-run. However, iteration may occur when the requestor needs an update of the search.

In the long run, it seems likely that a repeat customer will be one who has derived a measure of satisfaction from the retrieved material. This satisfaction may well be influenced by understanding the search process, thus prompting the requestor to participate in future iterations of the search process. We feel that requestor involvement is important, and that it can be encouraged partly by an understandable search history printout, which provides the means for evaluating the usefulness of terms that caused citations to be retrieved. The printed search history can serve as a very useful focal point for discussions between the requestor and the searcher about the search results. The experience of interpreting a search history in conjunction with its output should be helpful in the development of future searches.

The following recommendations are made with two goals in mind: understandability of output to the requestor, and through-put speed at the terminal (terminal productivity). We recommend keyboard procedures which will provide the requestor with a "readable" printed search history, and clearly indicate the terms and strategy which have produced the resulting citations. The recommended procedures are also relatively fast, though not the fastest possible procedures.

a. Initialization

DIALOG provides an initialization routine that is started by the BEGIN command (!). This routine prompts the searcher to keyboard the title, searcher, requestor, and mailing address information; this information is then printed at the top of the search history which accompanies any citations printed off-line. It is also printed and displayed at the terminal. Initialization results in a very useful and clearly identifiable search output record. However, the initialization routine presently requires a considerable amount of terminal time (an average of 3.0 minutes to initialize, according to the use statistics reported in an earlier section of this report).

One alternative to using the full initialization routine is to use BEGIN BYPASS. In this case the search history is not identified by requestor, searcher, or title.

Another alternative to using the full initialization routine for each search is to include several "questions" after one initialization. For this study we have used Lockheed's definition of "search" and "question" that is described in an earlier section of this report, i.e., a search is bounded by a BEGIN command and an END - BEGIN, disconnect, or END - disconnect combination. Questions within a search are bounded by additional END commands. Thus a BEGIN followed by searching commands followed by an END, more searching commands, an END and a new BEGIN, would be considered as one search with two questions. This was discussed in some detail in an earlier section of this report. When several questions are included after one initialization, the search strategy used for each question will be included in the search history printed off-line, but the citations printed off-line will usually correspond only to the latest section of the search history (the portion since the previous END command). Because the requestor's terms may be mixed in the sequence of searcher actions, the requestors will probably not be able to easily interpret the search history, if indeed they see it at all. However, terminal time may be saved by grouping several questions into one search, if some concepts or terms are used in more than one question. Such economies are probably most suitable for installations processing a large number of search requests, for this approach requires some experience on the part of the searchers.

We recommend that each logically distinct search or question be initialized separately. We feel that initialization provides quite considerable advantages, for subsequent handling and understanding of the printed output. It would be very useful if a quicker version of the initialization routine were provided, which would minimize the time disadvantage.

BEST COPY AVAILABLE

Initialization may be considerably speeded up in the present system by "stacking" responses, using the semicolon, and providing the initialization information before the questions are asked. By stacking we mean sending several commands during the same transmission burst. This may be achieved simply by keying each command or response to be sent, followed by a semicolon, another response and semicolon, and filling up to one line of display (62 characters) before pressing the INTERRUPT, RETURN, or CARRIAGE RETURN key. Figure 23 provides an example of both the regular, and the stacked method of initializing.

The stacking of searcher responses shown in Figure 23 accomplishes the whole initialization with a minimum of terminal wait time. Sending more than one line of display (62 characters) at one time, however, results in a truncation of the character string, and this can mean that semicolons to send subsequent commands will not be recognized. Stacking would be greatly facilitated if the length of command string that could be sent were extended (and perhaps made visible on the screen). Initialization would be facilitated if a several-line block were provided, without separate promptings, so that the outputs could be clearly identified with a smaller penalty in time.

b. Relative Merits of EXPAND-SELECT Combination vs. SELECT Alone

Although a search which is analyzed with the printed Thesaurus and term frequency lists and written out rather fully in advance may not benefit by use of the EXPAND command, we believe that the use of the EXPAND command is usually preferable to the use of the SELECT command by itself.

The EXPAND command works on two levels. The command EXPAND READING, or EREADING, results in a display of the alphabetically-nearest Descriptors and Identifiers surrounding the characters READING (see Figure 24). For each line the display provides a reference line number (E-number) on the left, the term, a postings figure on the right, and if the term is in the Thesaurus, a figure indicating the number of terms related to it. A second EXPAND may be used, e.g., EXPAND E6, to view those related terms. The resulting display (see Figure 25) is equivalent to the related terms (RT) listings under a given Descriptor in the printed Thesaurus. The related terms in the Thesaurus displays have R-numbers (R1-Rn) as reference numbers. They may be further EXPANDed; this is equivalent to looking from one (bold face) Descriptor heading to another in the printed Thesaurus.

Advantages of using the EXPAND command are:

-- Alphabetically-near Descriptors and Identifiers may be SELECTed quickly and easily from the display, using the E-number or a list of E-numbers which may include E-number ranges (e.g., SELECT E6-E10);

-- Related terms from the Thesaurus may be SELECTed easily from the display, using the R-numbers as described above for E-numbers;

-- The EXPAND command is not sensitive to typing mistakes; it generates a display surrounding whatever character string is given. If the typing mistake is near the end of the character string it may still

BEST COPY AVAILABLE

```
PLEASE ENTER THE FOLLOWING INFORMATION.
PRESS   INT   FOLLOWING EACH ENTRY.
-----------------------------------------
```

DIALOG prompts

```
SEARCH TITLE
                        TITLE-OF-SEARCH
SEARCHER NAME
                        MY-NAME
REQUESTOR NAME
                        REQUESTOR-NAME
MAIL ADDRESS
                        SHORT-ADDRESS
```

Searcher responds one line at a time

```
YOU MAY ACCESS THE FOLLOWING FILES :
  1*--ERIC THRU JUL 74       --ED*EJ
  2 --ERIC 1972 -- ED* EJ
  3 --CHEM ABST VOL 72 - 81 ISS 2
  4 --ERIC 1974 2 ND QTR --- ED
  5 --(*NOT ON-LINE*)
  6 --NTIS 1964-1974, ISSUE 15
  7 --CONTRA CDITA FILE
  8 --ENGR. INDEX 73-74 JULY
  9 --AIM & ARM 1973
 10 --NAT.AGR.LIB CAIN ISS JUNE 74
 11 --PSY. ABSTR. 1967-1973
 12 --INSPEC-PHYSICS ISS 15
 13 --INSPEC-ELECT. ENGINEERING ISS 15
 14 --INSPEC-COMPUTERS AND CONTROL ISS 1
 15 --ABI DATA BASE
 16 --CMA-EMA
 17 --PRED WEEKLY
 18 --F&S TEXT
ENTER NUMBER OF DESIRED FILE
                         1
```

DIALOG prompts
Searcher responds

```
TITLE          TITLE-OF-SEARCH
DATE/FILE      8-15-74 1
SEARCHER       MY-NAME
REQUESTOR      REQUESTOR-NAME
ADDRESS        SHORT-ADDRESS
            SET ITEMS DESCRIPTION
            --- ----- -----------
```

DIALOG reports back results of the initialization

Fig. 23a. Regular Method of Initializing
(Begin Command Not Shown)

```
                    [illegible]

               TITLE-OF-SEARCH [illegible]
PLEASE ENTER THE FOLLOWING INFORMATION.
PRESS 'INT' FOLLOWING EACH ENTRY.
------------------------------------------
SEARCH TITLE
SEARCHER NAME
REQUESTOR NAME
MAIL ADDRESS
YOU MAY ACCESS THE FOLLOWING FILES :
  1 --ERIC THRU [illegible]
  2 --ERIC 1972 [illegible]
  3 --CHEM ABST [illegible]
  4 --ERIC 1974 [illegible]
  5 [illegible]
  6 --NTIS 1964-1974, [illegible]
  7 --CONTRA COSTA FILE
  8 --[illegible]
  9 --AIM [illegible]
 10 --[illegible]
 11 --PSYCH. ABSTR. [illegible]
 12 --INSPEC-PHYSICS [illegible]
 13 --INSPEC-ELECT. ENGINEERING [illegible]
 14 --INSPEC-COMPUTERS AND CONTROL [illegible]
 15 --[illegible]
 16 --[illegible]
 17 --[illegible]
 18 --[illegible]
ENTER NUMBER OF DESIRED FILE
TITLE        TITLE-OF-SEARCH
DATE/FILE    8-11-74 1
SEARCHER     MY-NAME
REQUESTOR    REQUESTOR-NAME
ADDRESS      SHORT-ADDRESS
           SET ITEMS DESCRIPTION
           --- ----- -----------
```

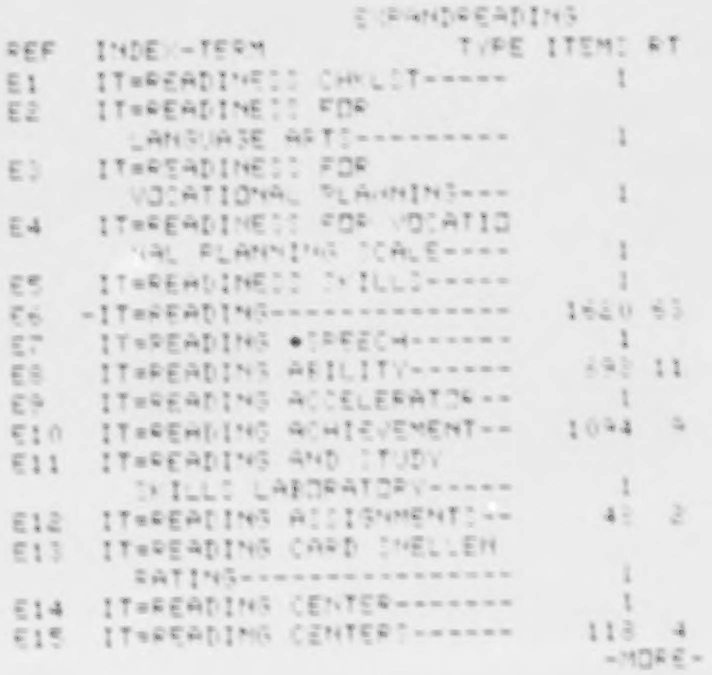
DIALOG prompts, but already [illegible] information

DIALOG repeats back results of the initialization

BEST COPY AVAILABLE

Fig. 23b. Stacked Method of Initializing

BEST COPY AVAILABLE



```
                    EXPANDREADING
REF   INDEX-TERM              TYPE ITEMS RT
E1    IT=READINESS CHKLIST-----        1
E2    IT=READINESS FOR
        LANGUAGE ARTS---------         1
E3    IT=READINESS FOR
        VOCATIONAL PLANNING---         1
E4    IT=READINESS FOR VOCATIO
        NAL PLANNING SCALE----         1
E5    IT=READINESS SKILLS-----         1
E6   -IT=READING---------------     1620 63
E7    IT=READING SPEECH-------         1
E8    IT=READING ABILITY------       692 11
E9    IT=READING ACCELERATOR--         1
E10   IT=READING ACHIEVEMENT--      1094  4
E11   IT=READING AND STUDY
        SKILLS LABORATORY-----         1
E12   IT=READING ASSIGNMENTS--        45  2
E13   IT=READING CARD SNELLEN
        RATING----------------         1
E14   IT=READING CENTER-------         1
E15   IT=READING CENTERS------       118  4
                                       -MORE-
```

Fig. 24. Sample EXPAND Display

```
                  [illegible]
REF   INDEX-TERM                  TYPE ITEMS  RT
[illegible]
[illegible]
[illegible]
[illegible]
[illegible]
[illegible]
[illegible]
[illegible]
[illegible]
[illegible]
[illegible]
[illegible]
[illegible]
[illegible]
[illegible]
[illegible]
[illegible]
[illegible]
[illegible]
[illegible]
                                           -MORE-

                  PAGE
REF   INDEX-TERM                  TYPE ITEMS  RT
[illegible]
[illegible]
[illegible]
[illegible]
[illegible]
[illegible]
[illegible]
[illegible]
[illegible]
[illegible]
[illegible]
[illegible]
[illegible]
[illegible]
[illegible]
[illegible]
[illegible]
[illegible]
[illegible]
[illegible]
                                           -MORE-
```

Fig. 25. Sample EXPAND Display for Related Terms

BEST COPY AVAILABLE

result in a useful display. (The SELECT command with a typing mistake in the term chosen will give back a zero-posting message and have to be repeated.)

-- Less information needs to be keyed in; e.g., keying in the first eight or ten characters of a term will bring a display in the general alphabetic area desired;

-- The EXPAND results in a lookup in a combined file of Descriptors and Identifiers, thus saving a double lookup in the printed reference tools;

-- The on-line indexes are usually more up to date than the printed Thesaurus and other printed searching aids.

A disadvantage is:

-- If related terms and alphabetically-near Identifiers are not used, the EXPAND-SELECT combination requires two commands to obtain one term.

c. Relative Merits of SELECT and Straight Typing vs. SELECT E- or R-numbers

When SELECTing terms, 5 options are available. These are:

Option 1: Terms may be SELECTed directly, by keyboarding the entire term: e.g.

```
SREADING PROGRAMS [INT]*

SREADING READINESS [INT]
```

Option 2: If an EXPAND has been used, E- or R-numbers can be SELECTed as follows:

```
SR1 [INT]

SR7 [INT]
```

When doing EXPAND - SELECT operations, some time can be saved by stacking commands in the same manner as in the initialization routine. That is, after viewing an EXPANDed display, SELECT several of the displayed terms before keying the INTERRUPT command (e.g., SELECT E6;SELECT E8;SELECT E10-E12 [INT]). Or call for the next EXPAND command along with the last SELECT command (e.g., SELECT E6;EXPAND ABSTRACTING [INT]). Remember that the stacked commands must not exceed 62 characters. On dial-up hard copy terminals with an 80 character line (where the printhead is positioned at the 21st character), stacked commands should all be contained on a single line of display before transmitting. On a CRT terminal the 62-character window may be more than one line long, depending on the width of the display screen.

---

* [INT] will be used to denote pressing the INTERRUPT, RETURN, CARRIAGE RETURN, or other send key. S is the form of the SELECT command we preferred, since it is one character and does not require use of the shift key. Other alternatives for the command are the # sign, or the full form, SELECT.

Options 3 and 4 are simply options 1 and 2, stacked.

Option 3: Entire terms may be keyboarded and stacked, e.g.

SREADING PROGRAMS; SREADING READINESS [INT]

With this option the size of the 62-character "window" for data entry usually precludes stacking more than two or three Descriptors. Also, if terminal errors occur, the effect of typing out full Descriptors, if they are lost by error, is noticed!

Option 4: After an EXPAND, several E- or R-numbers can be SELECTed by stacking, e.g.

SR1; SR3; SR9 [INT]

Options 1 through 4 result in each SELECTed term being displayed on the search history. This provides a clear record of exactly which terms have been used.

Option 5: Another mode of SELECTing involves a group of E- or R-numbers separated by commas (or hyphens), e.g.

SR1,R3-R5,R7 [INT]

With this option, the invididual terms are not shown on the search history. The E- or R-numbers are shown, along with the reference point to which they relate, e.g.

R1,R3-R5,R7

IT=READING

(Note: If many terms are selected, sometimes the reference point is truncated on the search history.)

Option 5 has the effect of creating an automatic ORed set. It is most efficient in terms of time, as long as changes are not required later in the search. The automatically ORed set, however, normally contains items from only one screen display, and thus unless all desired terms for the facet being developed are located on one display, a COMBINE command will still be needed to OR together the automatically ORed sets from several screen displays.

If it is decided later in the search that a certain term should not be included in the ORed set, the set will have to be created again, unless NOT logic is employed, with its attendant danger of excluding citations containing both acceptable and unacceptable Descriptors.

Option 5 implies limited requestor involvement in the search strategy, since the requestor cannot see, from a list of numbers and a reference point, what terms are actually being used (unless a hard copy terminal was used and the requestor is supplied with the full search record).

We recommend option 4 as the usual approach, a method which is fast, but which is also fully reported on the search history.

d. Timings of Five Alternative Procedures for SELECTing Descriptors

The 5 different methods of SELECTing Descriptors were timed, in order to obtain some idea of the time differences involved. Table 19 shows the results of this timing exercise.

SELECT straight typing and SELECT R-numbers were compared, using either individual sends ([INT] after each command), stacking (semicolon after each command), or chaining (comma after each R-number chosen). For the SELECT R-numbers section, a display ("EXPAND LIBRARIES") was generated before times were counted, and an END command was issued so that only the time for actually SELECTing would be measured. (The display of LIBRARIES and its related terms was not affected by the END command). We did not include the time used for EXPANDing to obtain the display, because the number of EXPAND commands preceding any one SELECT sequence could vary: one EXPAND would be needed to create an alphabetical E-number display from which items could be SELECTed; another EXPAND would be required to create a Thesaurus R-number display; a further EXPAND to look through the Thesaurus; perhaps a PAGE command to view a second page of a display. From the LIBRARIES display, ten terms were SELECTed -- the second, fourth, sixth, eighth, etc., related terms. After the ten had been selected, by whichever method, an END command was issued, and the elapsed time reported by the system was noted. This process was repeated five times for each of the five modes of SELECTing. Results indicate that option 5, "chaining" (using commas with one select command) is fastest, requiring an average of .34 minutes for ten non-adjacent terms from one screen. "Stacking" SELECT R-numbers (option 4) was next fastest, requiring an average of .74 minutes for selecting the same ten non-adjacent terms. (This is the option preferred by the project team.) Third fastest was option 3, the stacked sending of straight typed terms, with an average of 1.61 minutes to select the same 10 terms.

Individual sends for SELECT R-numbers (option 2) averaged 1.96 minutes for the 10 terms, while individual sends and straight typing (option 1) was the slowest method, averaging 2.23 minutes to SELECT the ten terms.

The difference between the extremes in time is almost two minutes for ten terms; this is enough time to consider seriously for routine procedures. (However, the overhead time of EXPANDing the term LIBRARIES in order to create the Thesaurus display would lessen this difference slightly. On the high-speed terminal at which this timing experiment was performed, the time required to create the display was minimal; however, on a slower terminal the time required to create the displays could be noticeably higher.

TABLE 19

RESULTS OF TIMING EXPERIMENT FOR SELECTING 10 DESCRIPTORS

	Time to SELECT 10 Descriptors (in Min.)				
	With Individual Send		With Stacked Send (;)		With Chained Send (,)
	Option 1	Option 2	Option 3	Option 4	Option 5
Trial Number	SELECT Straight Typing	SELECT R Numbers	SELECT Straight Typing	SELECT R Numbers	SELECT R Numbers
1	2.03	1.49	1.34	.71	.46
2	1.84	2.20	1.68	.89	.36
3	2.65	2.59	1.31	.75	.32
4	2.21	1.69	1.34	.65	.28
5	2.45	1.84	1.31	.70	.31
Average	2.23	1.96	1.61	.74	.34

Table 19 shows that stacking commands results in definite time savings. We recommend that stacking be used whenever possible. The difference in time between options 2 and 3 illustrates that efficiency in sending commands (use of stacking) has more effect on time used than efficiency in keyboarding (keying only short R-numbers rather than full Descriptors).

Between the two fastest methods, option 4 (SELECT R-numbers, stacked), and option 5 (SELECT R-numbers using commas, chaining) there was an average difference of only one-half minute for the ten terms. We believe that the use of option 4 is worth the extra half minute, since the resulting search history, displaying each term SELECTed, will have increased readability.

3. Strategy

a. Number of Terms Needed to Adequately Express Each Facet of a Multi-Facet Search

Due to the frequent use of broad concepts in searching the ERIC files, and to the characteristics of the ERIC indexing language, one of the questions faced by searchers at the terminal is the following: how many terms will adequately express a given facet (or aspect, or concept) of a multi-facet search? How much effort should be expended in looking for possible Identifiers (ERIC's free indexing terms), and for variant forms of terms, such as plurals, misspellings, and alternative punctuations?

The answer to these questions is not obvious when one is dealing with intersected sets. Sharon Jewell[15] has given some approximations of the kinds of retrieval quantities one may expect from intersecting heavily- and lightly-posted terms. To develop guidelines in terms of facets containing several terms, an attempt was made in the present study to measure the effect on retrieval of using varying numbers of terms to express each facet of some real questions, with two or three facets per question.

We measured the incremental effect (in number of output citations) of adding each additional term to the facets of 2- or 3-way intersected searches (searches incorporating 2 or 3 sets combined with AND). We chose to work from the heavily-posted terms outward, adding one term to each facet (unless the facet was a SEARCH SAVE) at each increment. We considered working backwards, by subtracting one term from each facet at each decrement, but rejected this method because the facets had widely divergent numbers of terms, and it would have been difficult to determine at what point to decrement the smaller facet.

As a source of real questions, search requests relevant to their personal interests were solicited from Ph.D. students in the University of California, Berkeley, School of Education, from ILR staff, and from some persons outside the University of California. The recipients of the searches agreed to make relevance judgments of the output.

Searches were negotiated during personal interviews, except for one search which was negotiated by a colleague of the requestor. The Thesaurus was used as a source of terms during these interviews, which were held before the searchers went to the terminal. High recall performance rather than precision, was emphasized in the formulation of the question.

At the terminal, the searches were first run in an "exhaustive" manner, trying to extract as many potentially relevant references as possible, in order to identify the set of relevant citations in the file. Variant forms (additional Descriptors, Identifiers, singular-plural forms and misspellings) were SELECTed whenever appropriate. A total of 14 searches, with a total number of 364 terms, was performed in this way.

We attempted to use as many terms as possible to express each facet in order to find the point of diminishing returns: i.e., the point at which the addition of further terms did not result in new citations being retrieved. For the searches done, facets were represented by a range of 1 to 19 terms with an average of 11.4 terms per facet. Each term was SELECTed separately in order to have it identified with its postings figure. This method was necessary for collecting the numerical data but also coincided with our choice of optimum searching approach as discussed in an earlier section.

In the searches used for this section of the study, two or three facets were used in one intersection. In some of the searches, several alternative set intersections were made; one was chosen for the study.

In most of the three-facet searches there was a facet representing age or grade level; such facets were handled by using established saved searches where they existed. Since SEARCH SAVE returns a complete, merged set of term postings, it was not meaningful to break the set apart in order to treat the individual terms incrementally. A searcher would not benefit by breaking apart sets returned by a SEARCH SAVE; if the sets were not appropriate the searcher should, at another time, create a separate saved search. For this reason we did not include SEARCH SAVE facets in the incremental treatment. One or two searches had grade level facets which were not already the subject of saved searches. These facets were incremented.

The searches were completed in the exhaustive manner, and results were sent to the users for relevance judgments. Relevance judgments were received in several different forms, ranging from a binary yes-no, through a "new" vs. "already seen" and "potentially useful" vs. "not worth looking into" judgment, to a 1-4 scale of relevance. From these judgments a binary rating was extracted, incorporating the "most relevant" citations, whether previously seen or new to the requestor.

Relevance is a judgment as to pertinence to an information need, eventually as perceived by the requestor of a search. Precision may be considered as the ratio of the relevant retrieved citations to the total set of retrieved citations. Recall may be considered as the ratio of relevant citations retrieved to the total set of relevant citations in the file (which is generally unknown). For an in-depth discussion of these and other measures, see Lancaster's Information Retrieval Systems: Characteristics, Testing and Evaluation.[16] Another excellent discussion may be found in King and Bryant's The Evaluation of Information Services and Products.[17]

Several of the searches achieved quite low precision figures. In some cases this is probably attributable to the difficulty of matching the search question to the data base; in some cases to insufficient experience on the part of the searchers, resulting in omission of important concepts which should have narrowed the search; in other cases it probably reflects the use to which the material was to be put. Ph.D. research is apt to be concerned with theoretical rather than practical aspects, and a good many of the retrieved citations reflected a "how to do it" approach which was not of interest to several of these users.

After the exhaustive run had been made, the searches were run again, this time selecting terms for each facet in decreasing order of the number of postings.

Then an iterative process began. Taking the most heavily posted term from each facet (except the grade-level facet which was usually expressed by a SEARCH SAVE, and was not changed in any way), we combined these using AND logic, and printed the resulting set (using format 1, for brevity), e.g.

```
C 1*11
```

In this and the following examples "C" indicates the COMBINE operation, "*" indicates the AND operation, while "+" indicates the OR operation.

Then we took the first two most heavily-posted terms from each facet, combined them, and printed the resulting set.

```
C (1+2)*(11+12)
```

This process was continued until the full set of citations retrieved by the original, exhaustive search was reached.

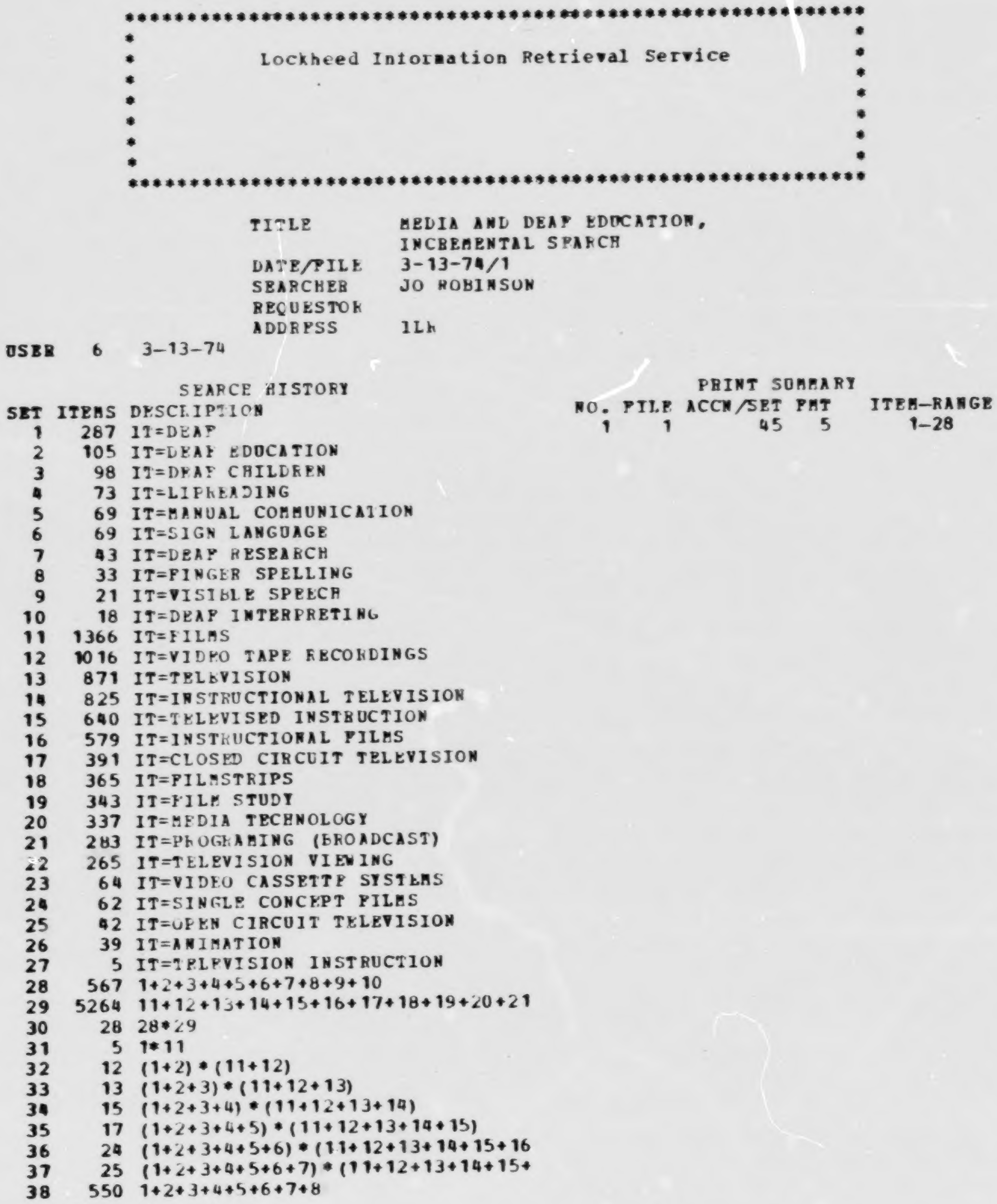
A three-facet search with one facet represented by a SEARCH SAVE was treated as follows:

```
C 1*11*21  (where set 21 represents a saved search)

C (1+2)*(11+12)*21

C (1+2+3)*(11+12+13)*21
```

An example of an incremental search is given in Figure 26.

Sixteen incremental searches were performed, and fourteen of them were used for this study. We obtained some very broad questions, and found that in these cases we were working very near to the limits set for the

BEST COPY AVAILABLE



```
***************************************************************
*                                                             *
*          Lockheed Information Retrieval Service             *
*                                                             *
*                                                             *
*                                                             *
*                                                             *
*                                                             *
***************************************************************

                  TITLE        MEDIA AND DEAF EDUCATION,
                               INCREMENTAL SEARCH
                  DATE/FILE    3-13-74/1
                  SEARCHER     JO ROBINSON
                  REQUESTOR
                  ADDRESS      1LR
USER   6   3-13-74

               SEARCH HISTORY                                PRINT SUMMARY
SET ITEMS DESCRIPTION                          NO. FILE ACCN/SET FMT    ITEM-RANGE
  1   287 IT=DEAF                                1    1       45   5       1-28
  2   105 IT=DEAF EDUCATION
  3    98 IT=DEAF CHILDREN
  4    73 IT=LIPREADING
  5    69 IT=MANUAL COMMUNICATION
  6    69 IT=SIGN LANGUAGE
  7    43 IT=DEAF RESEARCH
  8    33 IT=FINGER SPELLING
  9    21 IT=VISIBLE SPEECH
 10    18 IT=DEAF INTERPRETING
 11  1366 IT=FILMS
 12  1016 IT=VIDEO TAPE RECORDINGS
 13   871 IT=TELEVISION
 14   825 IT=INSTRUCTIONAL TELEVISION
 15   640 IT=TELEVISED INSTRUCTION
 16   579 IT=INSTRUCTIONAL FILMS
 17   391 IT=CLOSED CIRCUIT TELEVISION
 18   365 IT=FILMSTRIPS
 19   343 IT=FILM STUDY
 20   337 IT=MEDIA TECHNOLOGY
 21   283 IT=PROGRAMING (BROADCAST)
 22   265 IT=TELEVISION VIEWING
 23    64 IT=VIDEO CASSETTE SYSTEMS
 24    62 IT=SINGLE CONCEPT FILMS
 25    42 IT=OPEN CIRCUIT TELEVISION
 26    39 IT=ANIMATION
 27     5 IT=TELEVISION INSTRUCTION
 28   567 1+2+3+4+5+6+7+8+9+10
 29  5264 11+12+13+14+15+16+17+18+19+20+21
 30    28 28*29
 31     5 1*11
 32    12 (1+2)*(11+12)
 33    13 (1+2+3)*(11+12+13)
 34    15 (1+2+3+4)*(11+12+13+14)
 35    17 (1+2+3+4+5)*(11+12+13+14+15)
 36    24 (1+2+3+4+5+6)*(11+12+13+14+15+16
 37    25 (1+2+3+4+5+6+7)*(11+12+13+14+15+
 38   550 1+2+3+4+5+6+7+8
```

Figure 26. Example of an Incremental Search

DIALOG system, both in terms of number of sets used (DIALOG allows 98 sets) and in terms of number of postings involved in the sets created. In the incremental searches we were constantly creating new sets; if several of these sets had thousands of postings, disk space was used at a terrific rate; thus three of the broad searches came to an impasse with the message "DISK STORAGE OVERFLOW". One of the searches was nearly complete (it had achieved 97% of the retrieval from the exhaustive set and there was only one more term to increment), and is included in this study as if it were actually a completed search. The incremental searches for two others could not be completed because they ran into the system limitations.

Topics for the searches run are given in Appendix B.

Table 20 shows the incremental effect, in terms of the number of citations retrieved, of adding one more term to each facet. Remember that all of the terms have been added in order of decreasing frequency of postings. Terms added without increasing the amount of output are also shown. This same data is illustrated in Figure 27. Note that in eight out of 14 searches, from one to ten terms were used which had no incremental effect on the output retrieved, because the output had already been retrieved by other terms (with higher postings). Many of these terms which had no effect on the output retrieved were infrequently-used Identifiers, or had spelling, punctuation, or spacing errors. The results noted here are probably largely due to the redundancy and overlap of related terms used by ERIC indexers.

Table 20 and Figure 28 show the percentage of the total output citations (combined relevant and non-relevant) retrieved at each step. Using just the four most heavily-posted terms per facet, all but three searches had achieved more than 50% of the exhaustive output, and eight out of the 14 had achieved 70% or more of the full output. With ten terms per facet, over 96% of the citations had been retrieved in all 14 searches.

Table 20 and Figure 29 show the percentage of the relevant citations retrieved at each incremental step. Using just the four most heavily-posted terms, ten searches had achieved 50% or more of the output judged relevant. With ten terms per facet, all but one search had retrieved 92% or more of the output judged relevant, but two searches did not achieve the last relevant citation until the thirteenth term was reached. The fact that two searches required the thirteenth term (in rank order by number of postings) for completion of the set of relevant citations indicates that specific terms with low postings may sometimes be important to a search, and gives warning that searchers must not rely only on frequency of postings for information value. For a discussion of the inverse relationship between information value and frequency of term assignment, see Tell's The Use of ERIC Tapes in Scandinavia[18] and Williams' "Functions of a Man-Machine Interactive Retrieval System".[19]

From the results of this study we conclude that for ERIC searching, if exhaustivity is a requirement, there seems to be very little to be gained by using more than 10 of the most heavily-posted terms per facet. (Of course, some facets will be completely satisfied by less than 10 terms.)

BEST COPY AVAILABLE

TABLE 20

INCREMENTAL EFFECT OF ADDING ONE MORE TERM TO EACH FACET

Number of Terms In Search Facet	Deaf Children			Indian Reservations			Science Discovery			Joint Facilities			Institutional Service			Alternative Schools			Remedial Education			Environment Attitudes			Scientific Inquiry			Library Evaluation			Library Costs			Perceptual Salience			Pre-Service Programs			Bilingual Education		
	1*			2			3			4			5			6			7			8			9			10			11			12			13			14		
	NC	%C	%R	NC	%C	%R	NC	%C	%R	NC	%C	%R	NC	%C	%R	NC	%C	%R	NC	%C	%R	NC	%C	%R	NC	%C	%R	NC	%C	%R	NC	%C	%R	NC	%C	%R	NC	%C	%R	NC	%C	%R
1	5	18	10	21	60	45	0	0	0	2	7	0	10	21	28	11	17	0	34	39	50	10	8	15	26	18	14	9	4	1	9	4	2	1	1	0	253	76	78	28	7	7
2	12	43	45	28	80	64	14	30	31	7	24	40	11	23	28	36	55	29	43	49	50	70	55	63	78	54	38	79	36	33	48	20	6	33	13	6	288	87	89	64	16	17
3	13	46	50	32	91	82	18	39	38	19	66	60	12	25	28	41	63	43	54	62	61	93	73	81	134	93	94	103	47	47	107	43	42	44	18	6	302	91	94	134	33	40
4	15	54	55	33	94	91	22	48	38	20	69	70	20	42	44	53	82	86	60	69	78	108	84	93	144	100	100	112	51	52	172	70	71	125	50	29	314	95	100	170	42	40
5	17	61	65	34	97	100	29	63	62	23	79	70	23	48	47	57	88	93	76	87	94	115	90	93	144	100	100	125	57	60	190	77	75	139	56	41	321	97	100	211	52	43
6	24	86	95	34	97	100	33	72	69	25	86	70	35	73	84	65	100	100	79	91	94	116	91	96	144	100	100	131	60	61	202	82	76	156	63	52	324	98	100	236	58	45
7	25	89	95	34	97	100	40	87	77	26	90	70	36	75	88				85	98	100	117	91	96	144	100	100	196	90	88	217	88	76	187	75	52	332	100	100	275	68	48
8	27	96	95	34	97	100	43	93	92	26	90	70	45	94	94				87	100	100	117	91	96	144	100	100	208	95	95	241	98	90	195	79	71	332	100	100	279	69	50
9	27	96	95	34	97	100	45	98	92	26	90	100	45	94	94				87	100	100	128	100	100	144	100	100	209	96	97	246	100	100	216	87	71	332	100	100	288	71	52
10	27	96	95	34	97	100	45	98	92	26	90	100	47	94	97				87	100	100	128	100	100				214	98	97	246	100	100	223	90	82	332	100	100	375	93	98
11	27	96	95	35	100	100	46	100	100	26	90	100	47	94	97				87	100	100	128	100	100				216	99	98	246	100	100	241	97	94	332	100	100	384	95	100
12	28	100	100	35	100	100				28	97	100	47	94	97				87	100	100	128	100	100				217	99	99	246	100	100	248	100	100	332	100	100	405	100	100
13	28	100	100	35	100	100				29	100	100	48	100	100							128	100	100				218	100	100	246	100	100				332	100	100	405	100	100
14	28	100	100	35	100	100							48	100	100							128	100	100				218	100	100	246	100	100				332	100	100	405	100	100
15	28	100	100	35	100	100							48	100	100							128	100	100							246	100	100									
16	28	100	100																												246	100	100									
17	28	100	100																												246	100	100									
18																															246	100	100									
19																															246	100	100									

* Search Number

NC = Number of citations retrieved

%C = Percent of total citations retrieved at each step

%R = Percent of the citations judged relevant retrieved at each step

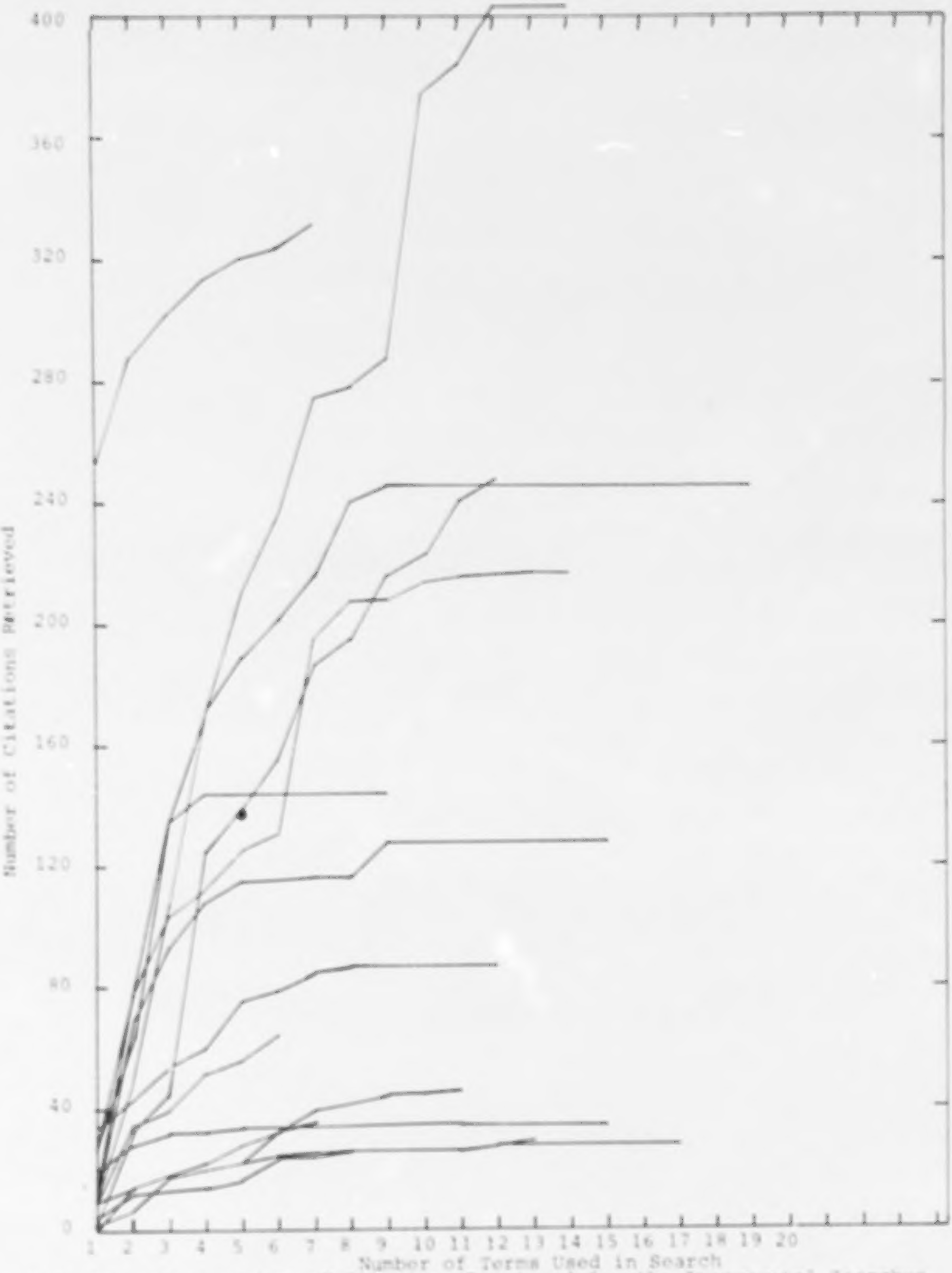


Fig. 27. Total Number of Citations Retrieved for the Incremental Searches

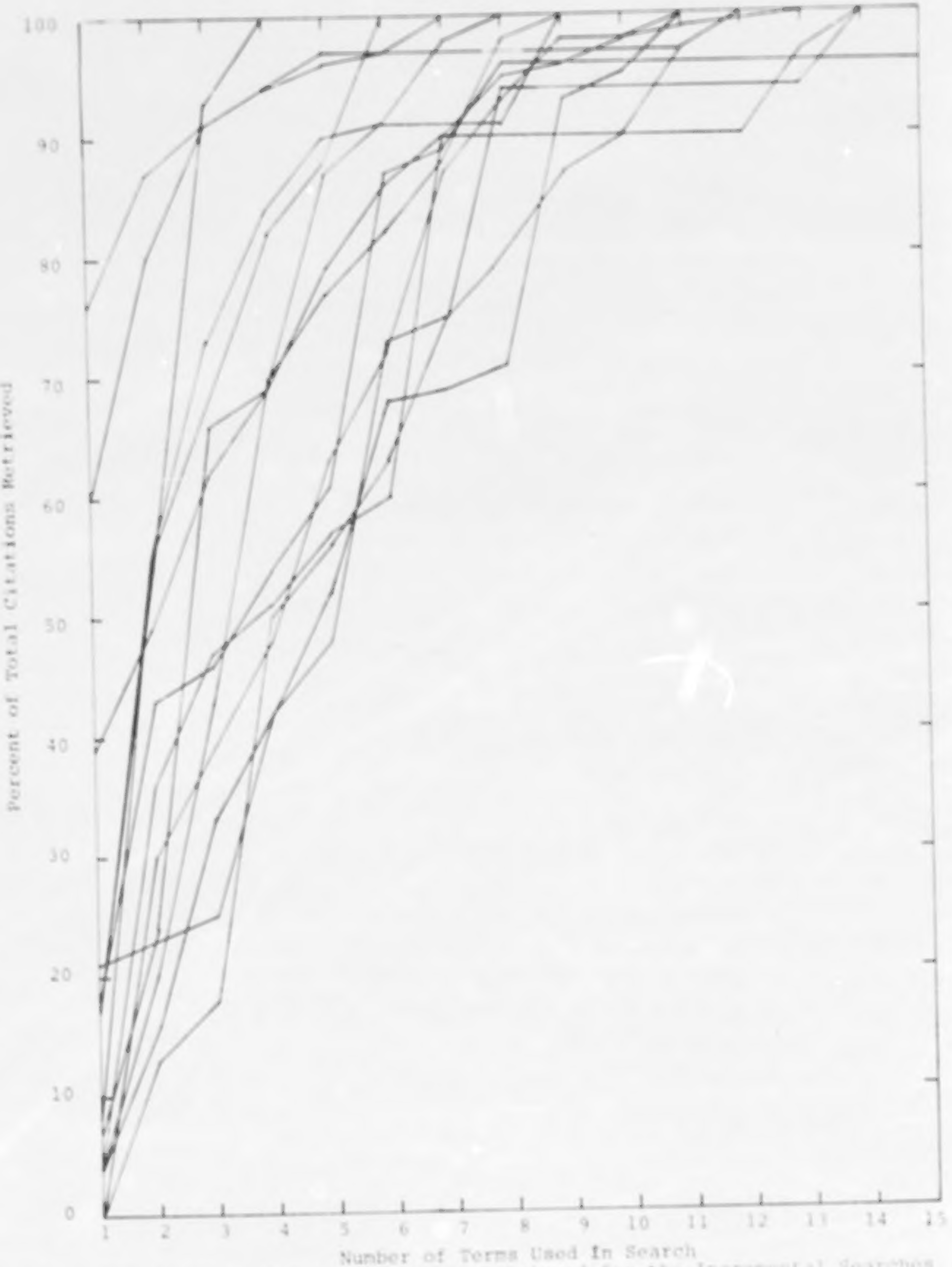
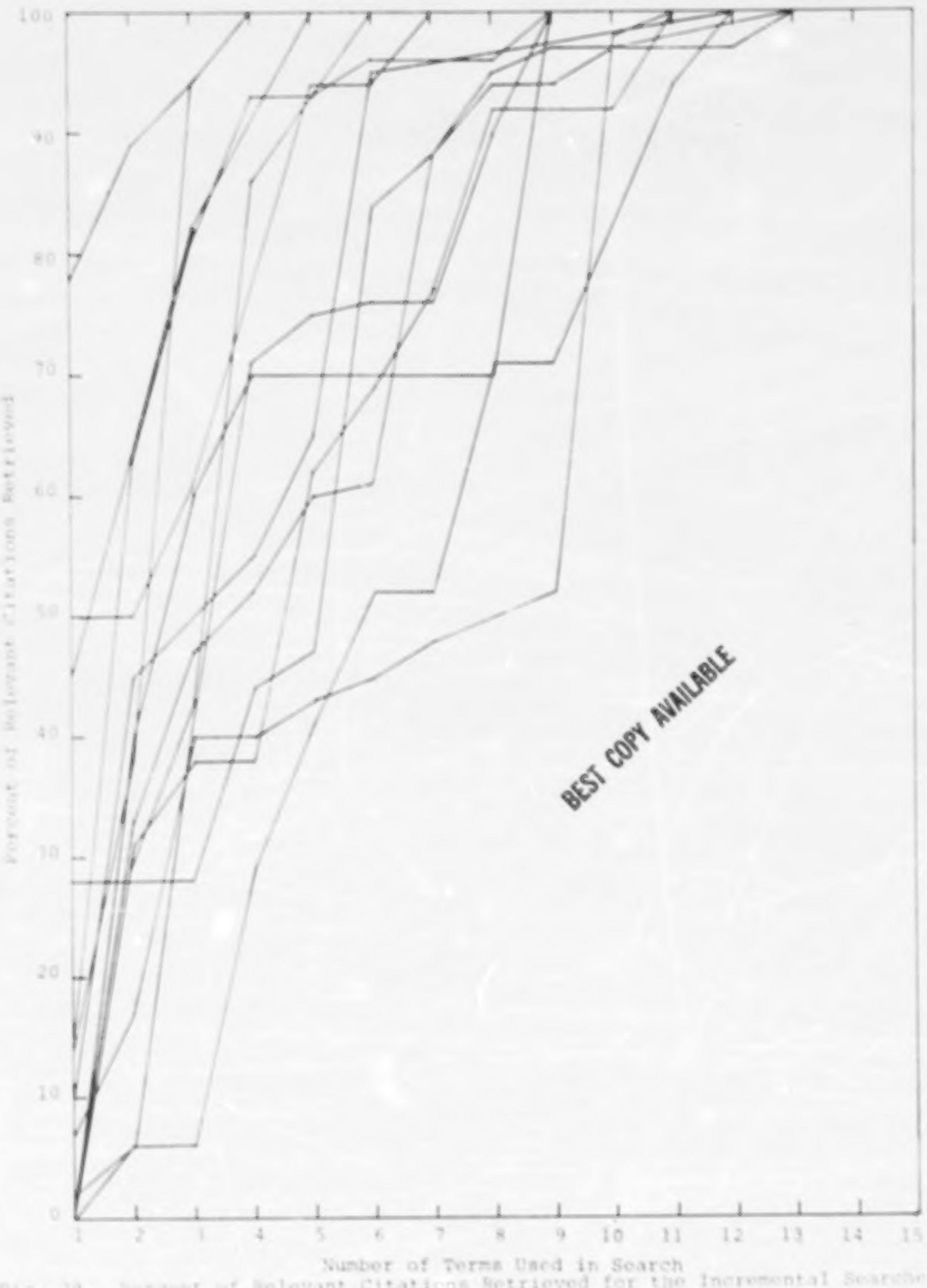


Fig. 28. Percent of Total Citations Retrieved for the Incremental Searches



BEST COPY AVAILABLE

Fig. 29. Percent of Relevant Citations Retrieved for the Incremental Searches

For fairly exhaustive searches with two facets, then, it is possible that up to 20 terms would be needed, while for three facets up to 30 terms would be needed to retrieve most of the relevant citations. These are, of course, merely guidelines, and not absolute figures.

This does *not* mean that relevant terms in excess of 10 should be routinely ignored, nor that selection of terms should be made solely on the basis of number of postings. Indeed, the choice of terms *must* be made by meaning, not by number of postings.

Our study does indicate, however, that the extra effort of searching for misspellings, or peripherally-relevant terms is not productive. This finding agrees with the data on variant forms that was reported in an earlier section of this report. Centrally relevant Descriptors or Identifiers, regardless of their number of postings, should certainly be included in the search statement.

Terms with very few postings will seldom have an impact if the central Thesaurus terms appropriate to the search are used. Terms with few postings should probably only be chosen if they are specifically pertinent to the search topic.

b. Methods of Limiting Quantity of Output and Their Effect on Relevance

Near the end of an on-line search on a given question, one is sometimes confronted with a "too large" set of output citations. Aside from changes in the search facets or terms from a subject point of view, the general procedure followed to reduce the size of such final output sets is to use one of the LIMIT options available with DIALOG. In an earlier section of this report, the analysis of frequency of use of commands by terminal operators shows that the LIMIT command is used 0.84 times per question, accounting for 3.40% of the commands used per question.

In an effort to identify factors that might be helpful in limiting the selected output, we investigated the effect that several different kinds of limiting factors would have on the number of citations retrieved, and on the percentage of relevant citations (from the exhaustive set) retrieved.

(1) Limit by LIMIT/MAJ Command

In the ERIC system, major applicability of a given Descriptor or Identifier to a given document is indicated by the presence of an asterisk preceding that term. In DIALOG, major value is represented by the MAJOR sub-command of the LIMIT command. The Summary of Significant Rules in the *ERIC Processing Manual*[13] specifies that "Major Descriptors (identified by a preceding asterisk) are limited to five (5) per document. The maximum number of Descriptors is not limited but will depend on the nature of the document. Major Identifiers are limited to one (1) per document. The maximum number of Identifiers is not limited." In practice, 10 to 12 Descriptors are typically used in ERIC indexing, while one to three Identifiers may be present.

BEST COPY AVAILABLE

The LIMIT/MAJ command allows the searcher to reduce an output set by specifying that it must contain only Descriptors or Identifiers designated by the ERIC indexers as having major value. As approximately half of the terms for a given citation may be major terms, the yield for a given facet could be expected to be cut in half by use of the LIMIT/MAJ command. An intersected set could be expected to be greatly reduced by this method.

The LIMIT/MAJ command returns a set satisfied only by Descriptors or Identifiers marked by an asterisk (ERIC's way of indicating major applicability of this Descriptor to this document).

We were interested to see whether the use of the LIMIT/MAJ command resulted in an output containing relatively more of the relevant citations (higher precision than the exhaustive set), or whether a decrease in output would be coupled with a proportionate decrease in the number of relevant citations (same precision).

When used on an intersected set, the LIMIT/MAJ command, as implemented at the time of the study, required at least one term in each intersected facet to have MAJOR value (i.e., to carry an asterisk). In a three-way intersection, this can be very restrictive because some of the facets may not include a major term.

The Lockheed DIALOG Terminal Users Reference Manual[20] suggests, "Use the limit command on key conceptual terms so that only major Descriptors will be selected. It is usually best to limit individual terms rather than sets resulting from the combination of terms."

We investigated the LIMIT/MAJ command along these lines, using this command on the ORed sets of terms making up the "key" concepts, rather than on "key conceptual terms" alone. However, in some searches it was difficult to decide which concept should be considered the key concept, so we used the LIMIT/MAJ command separately on each facet except the grade level. This parallels the treatment of the grade level facet in the incremental searches.

This task was carried out at the same time as the incremental searches described in an earlier section of this report. Having obtained the exhaustive sets for each of the 14 searches, we proceeded to use DIALOG's LIMIT command and other methods to cut down on the quantity of output.

Eleven of the 14 searches described in the earlier section were used for this study. Of the three searches not included, one triggered "DISK STORAGE OVERFLOW", and because of a searcher error the output sets for two were not printed.

The LIMIT/MAJ command was used on each facet except the grade level facet. Each "MAJOR" facet was then intersected with its partner facet in unlimited form, e.g., the partner set could contain asterisked or non-asterisked (MAJOR or MINOR) terms. The results of these intersections were ORed together, giving a set where a MAJOR term from either but not necessarily both sets was present. This ORed set corresponds to the way the LIMIT/MAJ command operated on an intersected set in an earlier version of DIALOG.

These sets were then compared with each other and with the set resulting from using the LIMIT/MAJ command on the previously intersected set, in terms of relevant citations retrieved. The percentage of the relevant citations from the exhaustive set was determined.

An example of the way in which sets were limited and combined is given in Figure 30. The results of this experimental work are described in a later section of this report.

(2) Limit by Accession Number

Although DIALOG implements the accession number range as one parameter of the LIMIT command, for expediency this operation was done manually from the output of the exhaustive set. We feel that, whereas the ability to limit by time is very important, using an accession number range as presently required under LIMIT command is an awkward means by which to achieve this effect. One must look up the desired accession numbers in a table printed for the ERIC *Chronolog*; one must also LIMIT ED numbers and EJ numbers separately, because they are separate series of numbers. It would be far easier to use a year-month parameter for accession range limitation; such a parameter would have the additional great advantage of being useful across data bases.

In fact, RIE issue numbers are already stored as an inverted file (Figure 31). If this file were made numerical and stored as YYMM it would be much more useful.

It would be very helpful if Lockheed would implement a YYMM chronological feature as one of the options available with the LIMIT command, independent of the data base being searched. The accession number prefix (ED, EJ) could still be accepted as a modifier when appropriate.

(3) Limit by Printing Only the First N Citations

Several installations have used this type of output limiting; we tested how this method affected the number of relevant citations. Seven of our 14 searches produced more than 100 citations; using a limit of 100 citations the percentage of relevant citations was calculated for six of these searches.

Three researchers for whom searches were done during this study mentioned finding more "good" citations towards the front of the output. This may reflect greater timeliness (the newest citations are always printed first, unless the searcher requests another sorting sequence), or better acquisition efforts by the ERIC clearinghouses in more recent times, or greater applicability of CIJE citations (which appear at the front of the output) to their research.

(4) Results of All Attempts to Limit Output

Using the LIMIT/MAJ command on a previously intersected set, an average of only 33% of the relevant citations was retrieved from an average of 27% of the total citations. This seems to be a too-radical solution to the problem of too much output.

BEST COPY AVAILABLE

```
MEDIA AND DEAF EDUCATION

                SEARCH HISTORY
SET  ITEMS DESCRIPTION
  1    287 IT=DEAF
  2    105 IT=DEAF EDUCATION
  3     98 IT=DEAF CHILDREN
  4     73 IT=LIPREADING
  5     69 IT=MANUAL COMMUNICATION
  6     69 IT=SIGN LANGUAGE
  7     43 IT=DEAF RESEARCH
  8     33 IT=FINGER SPELLING
  9     21 IT=VISIBLE SPEECH
 10     18 IT=DEAF INTERPRETING
 11   1366 IT=FILMS
 12   1016 IT=VIDEO TAPE RECORDINGS
 13    871 IT=TELEVISION
 14    825 IT=INSTRUCTIONAL TELEVISION
 15    640 IT=TELEVISED INSTRUCTION
 16    579 IT=INSTRUCTIONAL FILMS
 17    391 IT=CLOSED CIRCUIT TELEVISION
 18    365 IT=FILMSTRIPS
 19    343 IT=FILM STUDY
 20    337 IT=MEDIA TECHNOLOGY
 21    283 IT=PROGRAMING (BROADCAST)
 22    265 IT=TELEVISION VIEWING
 23     64 IT=VIDEO CASSETTE SYSTEMS
 24     62 IT=SINGLE CONCEPT FILMS
 25     42 IT=OPEN CIRCUIT TELEVISION
 26     39 IT=ANIMATION
 27      5 IT=TELEVISION INSTRUCTION
 28    567 1+2+3+4+5+6+7+8+9+10
 29   5264 11+12+13+14+15+16+17+18+19+20+21
 30     28 28*29


 46    311 28/MAJ
 47   3285 29/MAJ
 48      8 30/MAJ
 49     12 46*29


 53    .13 47*28
 54     17 53+49
```

LIMIT/MAJ

Figure 30. Example of Use of LIMIT/MAJ on Different Set Combinations

BEST COPY AVAILABLE

```
         EXPAND IS=CIJE
REF    INDEX-TERM                         TYPE ITEMS RT
E1     IS=RIEMAR69----------------              966
E2     IS=RIEMAR70----------------              525
E3     IS=RIEMAR71----------------              745
E4     IS=RIEMAR72----------------             1015
E5     IS=RIEMAR73----------------             1199
E6     IS=RIEMAY69----------------              339
E7     IS=RIEMAY70----------------              372
E8     IS=RIEMAY71----------------             1502
E9     IS=RIEMAY72----------------              934
E10    IS=RIEMAY73----------------             1413
E11    IS=RIENOV69----------------              759
E12    IS=RIENOV70----------------              874
E13    IS=RIENOV71----------------              947
E14    IS=RIENOV72----------------             1265
E15    IS=RIENOV73----------------             1333
E16    IS=RIEOCT69----------------              345
E17    IS=RIEOCT70----------------              933
E18    IS=RIEOCT71----------------              959
E19    IS=RIEOCT72----------------             1507
E20    IS=RIEOCT73----------------             1375
```

Fig. 31. Example of RIE Issue Numbers Stored as an Inverted File

The results of LIMITing one facet to MAJOR and intersecting the resultant set with its partner set in full form produces better results in most cases. An average of 60% of the relevant citations was retrieved from an average of 53% of the total citations. However, the best use of LIMIT/MAJ for these searches appears to be the ORed set of intersections of one MAJOR facet with a partner non-major facet, and vice versa. In the study, this method retrieved an average of 85% of the relevant citations, from an average of 75% of the total citations. In nine of the 11 searches studied, this ORed set retrieved a higher percentage of the relevant citations than of total citations. None of the other possible ways to use LIMIT/MAJ appears to be as successful. It must be noted, however, that the ORed set we are speaking of does not reduce the total number of citations retrieved as effectively as some of the other versions.

We feel that this data indicates that it would be helpful if another type of LIMIT/MAJ command were implemented, which would retrieve items having a MAJOR posting in either (any) facet of an intersected set. This command should not replace the present LIMIT command, but should simply provide another option. A further option which would be very useful in cases where requestors want output only from a given time period would be the provision of a LIMIT/YYMM-YYMM feature. This could augment or replace the present LIMIT by accession number feature.

As to recommendations to searchers, no hard and fast rules can be given. We have attempted to provide a "menu" of possible methods of reducing output quantity, and must leave it to the discretion of the individual searchers which method they choose.

D. REFERENCES

1. Charles Missar, Input and Output Limitations. How They Can Affect On-Line Search Results. National Institute of Education. Unpublished. May 1973.
2. DIALOG Terminal Users Reference Manual. Palo Alto, Calif.: Lockheed Palo Alto Research Laboratory. Lockheed Retrieval Service Information Systems Laboratory. n.d., var. paging.
3. Interchange. A newsletter prepared by the staff of the ERIC Processing and Reference Facility. Bethesda, Md. Ceased publication as a formal numbered serial with the March 1974 issue. To continue on an irregular basis as information of interest and value to ERIC users is accumulated.
4. DIALOG Chronolog. A newsletter prepared by Lockheed Information Systems, Lockheed Palo Alto Research Laboratory, Palo Alto, Calif.
5. ERIC Descriptor and Identifier Usage Report. New York: Macmillan Information, 1973. 1533 p.
6. Title Index. Bethesda, Md. Leasco Systems and Research Corporation.
7. Report/Project Number Index. Bethesda, Md. Leasco Systems and Research Corporation.
8. Contract/Grant Number Index. Bethesda, Md. Leasco Systems and Research Corporation.
9. Clearinghouse Number to ED Number Cross Reference List. Bethesda, Md. Leasco Systems and Research Corporation.
10. Thesaurus of ERIC Descriptors. New York: CCM Information Corp., 1974. (New editions published annually).
11. Descriptor and Identifier Usage in RIE and CIJE. Produced by Lockheed Information Systems. ERIC/DIALOG Retrieval Services. Bethesda, Md.: ERIC Processing and Reference Facility. August 1974. 142 p.
12. Dictionary for ERIC/CIJE File Thru Sept. 1974, and Dictionary for ERIC/RIE File Thru Sept. 1974. Research Triangle Park, NC: North Carolina Science and Technology Research Center, 1974. 3 microfiche.
13. ERIC Processing Manual. Rules and guidelines for the acquisition, selection, and technical processing of documents and journal articles by the various components of the ERIC network. Bethesda, Md.: ERIC Processing and Reference Facility. July 1974. 544 p. ED-092 164.
14. ERIC Search Saves. Palo Alto, Calif.: Lockheed Missiles and Space Co., Inc. Lockheed Information Systems. 22 Oct. 1974.

15. Jewell, Sharon, and W. T. Brandhorst. *Search Strategy Tutorial; Searcher's Kit*. Washington, D.C.: National Institute of Education, October 1973. 86 p. ED-082 763.

16. Lancaster, F. Wilfrid. *Information Retrieval Systems. Characteristics, Testing and Evaluation*. New York: Wiley, 1968. p. 54-63.

17. King, Donald W. and Edward C. Bryant. *The Evaluation of Information Services and Products*. Washington, D.C.: Information Resources Press, 1971. p. 20-33.

18. Tell, Bjorn V., et al. *The Use of ERIC Tapes in Scandinavia; Searching with Thesaurus Terms in Natural Language*. Strasbourg, France: Council of Europe; and Stockholm, Sweden: Council for Cultural Cooperation; Royal Institute of Technology, 11 November 1972. 23 p. (ECS-DCC-72-15). ED-072 794.

19. Williams, J. H., Jr. "Functions of a Man-Machine Interactive Information Retrieval System" *Journal of the American Society for Information Science* 22:5 (September-October 1971) 311-317.

20. *DIALOG Terminal Users Reference Manual*. Appendix A. p. A11-A14.

BEST COPY AVAILABLE

APPENDIX A

Table 1. Chronological Record of Searches Performed by All Searchers, October 18 through November 1.

Searcher	Day	Time at Start of Search	Time at End of Search	Elapsed Time Reported by DIALOG (Minutes)‡
JT	Thurs. Oct. 18	5:18	5:35	15.98
AH	"	5:36	5:55	15.56
AH	"	5:55	6:11	16.23
JT	"	6:14	6:30	14.44
JT	"	6:36	6:45	14.39
AH	"	6:48	7:06	17.16
AH	"	7:07	7:26	18.03
JT	"	7:29	7:45	14.87
JT	"	8:13	8:29	15.08
AH	"	8:30	8:51	19.66
JT	"	8:52	9:10	16.83
JR	"	9:17	9:35	16.00
JT	"	9:32	9:50	15.80
AH	"	9:50	10.08	17.19
JR	"	10.25	11:15	31.43*
JT	"	11:15	11:34	16.56
JR	"	11:45	12:00	17.35
JT	"	11:55	12:16	19.58
AH	"	12:18	12:35	16.54
AH	"	12:35	12:50	12.44
JT	"	12:55	1:06	12.18
JT	"	1:08	1:20	12.12
JT	Thurs. Oct. 25	5:25	5:38	13.41
JT	"	5:40	5:55	14.05
DT	"	5:55	6:19	25.17
DT	"	6:17	6:41	17.12
JT	"	6:50	7:02	13.28
JT	"	7:32	7:45	12.06
DT	"	8:20	8:43	16.90
JT	"	9:40	9:55	13.35
JT	"	10:00	10:15	13.33
AH	"	10:15	10:39	20.86
AH	"	10:40	10:56	15.12
JT	"	11:01	11:16	14.87
JT	"	11:36	12:15	13.92*
AH	"	12:18	12:36	16.06
AH	"	12:36	12:51	14.46
AH	"	12:51	1:06	14.29
JR	"	1:10	1:30	18.81

*Indicates that the system was down during part of this search.

‡The elapsed time reported by DIALOG on-line at the end of each search.

Table 1 (continued)

Searcher	Day	Time at Start of Search	Time at End of Search	Elapsed Time Reported by DIALOG (Minutes)
JT	Fri. Oct. 26	5:10	5:22	11.02
JT	"	5:25	6:36	11.88
AH	"	5:36	5:50	13.28
AH	"	5:50	6:08	16.05
JT	"	6:10	6:25	13.94
JT	"	6:26	6:40	12.37
AH	"	6:38	6:52	12.97
JT	"	6:56	7:10	13.09
JT	"	7:11	7:25	13.18
AH	"	7:25	7:39	13.54
JT	"	8:05	8:45	25.61*
AH	"	8:52	9:08	15.36
JR	"	9:16	9:34	16.74
JR	"	9:34	9:50	13.21
JR	"	9:51	10:07	16.31
JT	"	10:10	10:36	14.37
JT	"	10:30	10:49	15.62
JT	"	10:50	11:08	17.14
AH	"	11:08	11:31	21.68
AH	"	11:34	11:57	18.07
JR	"	12:01	12:20	18.24
JT	"	12:25	12:40	14.12
JT	"	12:45	12:59	13.44
JT	"	1:00	1:15	16.27
JR	Mon. Oct. 29	5:15	5:27	10.51
JR	"	5:27	5:40	11:01
JR	"	5:40	5:52	11:67
JR	"	5:53	6:10	14.52
JR	"	6:12	6:24	13.34
JR	"	6:30	6:44	12.59
JR	"	6:44	6:58	12.89
JR	"	6:58	7:13	12.23
JR	"	7:14	7:26	13.76
JR	"	8:07	8:21	11.65
JR	"	8:22	8:35	12.16
JR	"	8:35	8:49	13.24
JR	"	8:50	9:03	13.10
JR	"	9:05	9:17	11.59
JR	"	9:24	9:39	14.87
JT	"	9:45	10:16	14.80*
JT	"	10:16	10:31	13.71
JT	"	10:32	10:58	23.29*
JT	"	11:00	11:12	12.47
JR	"	11:28	11:45	14.72
JT	"	11:58	12:10	12.70
JT	"	12:20	12:34	13.55
JT	"	12:35	12:50	15.04
JT	"	12:51	1:10	18.37
JT	"	1:10	1:25	12.33

Table 1. (continued)

Searcher	Day	Time at Start of Search	Time at End of Search	Elapsed Time Reported by DIALOG (Minutes)
DT	Tues. Oct. 30	5:00	5:15	14.96
DT	"	5:16	5.31	14.90
DT	"	5:34	5:55	13.03
DT	"	5:55	6:08	13.49
DT	"	6:10	6:22	12.17
DT	"	6:23	6:37	12.32
DT	"	6:39	6:52	12.43
AH	"	6:54	7:09	14.27
AH	"	7:10	7:24	13.09
DT	"	7:25	7:38	12.08
AH	"	8:01	8:14	11.93
AH	"	8:15	8:30	13.83
DT	"	8:32	8:44	11.73
DT	"	8:45	8:56	10.48
DT	"	8:56	9:02	14.07
JR	"	9:12	9:26	12.30
JR	"	9:43	10:02	11.82
JR	"	10:03	10:14	10.87
JR	"	10:37	11:45	15.78
AH	"	11:52	12:12	18.46
AH	"	12:13	12:32	18.29
DT	"	12:33	12:54	19.49
DT	"	12:59	1:17	16.60
DT	Wed. Oct. 31	5:39	5:55	15.09
DT	"	6:01	6:17	15.37
DT	"	6:23	6:40	14.81
DT	"	6:55	7:15	16.18
AH	"	7:20	7:37	16.38
AH	"	10:01	10:24	20.57
AH	"	10:25	10:45	18.79
DT	"	10:45	11:00	14.25
DT	"	11:00	11:25	16.97
JR	"	11:17	11:32	13.49
JR	"	11:33	11:50	15.28
JR	"	11:50	12:05	13.97
JR	"	12:06	12:28	18.90
JR & DT	"	12:42	1:12	21.89
DT	"	1:16	1:31	14.34
DT	"	1:32	1:43	11.69
JR	Thurs. Nov. 1	5:14	5:30	14.28
JR	"	5:30	5:43	12.21
JT	"	5:46	6:00	12.50
JR	"	6:09	6:25	15.04
JR	"	6:25	6:40	13.84
JT	"	6:45	7:00	14.99
JT	"	7:05	7:18	13.35
JR	"	7:20	7:35	15.23
JR	"	8:00	8:13	12.44
JT	"	8:15	8:30	13.40
JT	"	8:30	8:45	12.94
JT	"	8:45	9:00	14.80
DT	"	9:20	9:34	13.12
DT	"	9:35	9:47	11.14
DT	"	9:47	10:00	13.17
DT	"	10:01	10:13	10.92

Table 1. (continued)

Searcher	Day	Time at Start of Search	Time at End of Search	Elapsed Time Reported by DIALOG (Minutes)
JT	Thurs. Nov. 1	10:55	11:08	12.33
JT	"	11:08	11:20	11.87
JT	"	11:20	11:40	15.93*
JR	"	11:40	11:57	14.27
JT	"	11:58	12:14	16.77
JT	"	12:42	12:56	13.70
JR	"	12:58	1:12	13.12
JT	"	1:15	1:28	12.80

Table 2. Elapsed Time for Each Search Arranged According to Searcher (Excluding Searches Done When System Was Down), with Cumulative Mean for Each Searcher.

Searcher:	AH		JT		JR		JT	
	Elapsed Times	Cumulative Mean	Elapsed Times	Cumulative Mean	Elapsed Times	Cumulative Mean	Elapsed Times	Cumulative Mean
Thurs., Oct. 18	15.56	15.56	15.98	15.98	16.00	16.00		
	16.23	15.90	14.44	15.21	17.35	16.68		
	17.16	16.32	14.39	14.94				
	18.03	16.75	14.07	14.92				
	19.66	17.33	16.88	15.31				
	17.19	17.31	15.80	15.39				
	16.54	17.20	16.56	15.56				
	12.44	16.60	19.58	16.06				
			12.18	15.63				
			12.12	15.28				
Mean Elapsed Time:	16.60		15.28		16.68			
Thurs., Oct. 25	20.86	17.07	13.41	15.11	18.81	17.39	25.17	25.17
	15.12	16.88	14.05	15.02			17.12	21.15
	16.06	16.80	13.28	14.89			16.90	19.73
	14.46	16.61	12.06	14.69				
	14.29	16.43	13.85	14.63				
			13.83	14.58				
			14.87	14.60				
Mean Elapsed Time:	16.16		13.62		18.81		19.73	
Friday Oct. 26	13.28	16.21	11.02	14.40	16.74	17.23		
	16.05	16.20	11.88	14.26	13.21	16.42		
	12.97	15.99	13.94	14.25	16.31	16.40		
	13.54	15.85	12.37	14.16	18.24	16.67		
	15.36	15.82	13.09	14.11				
	21.68	16.13	13.18	14.07				
	18.07	16.23	14.37	14.08				
			15.62	14.14				
			17.14	14.26				
			14.12	14.25				
			13.44	14.22				
			16.27	14.30				
Mean Elapsed Time:	15.85		13.87		16.15			

BEST COPY AVAILABLE

Table 2. (Continued)

Searcher:	[illegible]		[illegible]		[illegible]		[illegible]	
Mon, Oct. 29	Elapsed Times	Cumulative Mean	Elapsed Times	Cumulative Mean	Elapsed Times	Cumulative Mean	Elapsed Times	Cumulative Mean
			13.71	14,28	13,51	15.94		
			12,47	11,[illegible]1	15.35			
			12,70	14,17	11,67	14,99		
			13.55	14,15	14.52	14,94		
			15.04	14,18	13,34	14,81		
			13,37	14,30	12,59	14,64		
			12,33	14,24	12,89	14,51		
					12,23	14,36		
					13,76	14,32		
					11,65	14,17		
					12,16	14,06		
					13,24	14,01		
					13,10	13,97		
					11,59	13,85		
					14,07	13,92		
					14,72	13,94		
Mean Elapsed Time:			14,12		12,74			
Tues, Oct. 30	14,27	16,13			12,30	13,67	14,96	16,54
	13,39	16,[illegible]			11,82	13,79	14,90	17,01
	11,93	15,[illegible]			10,87	13,67	13,03	17,01
	13,83	15,74			15,78	13,75	13,49	16,51
	15,46	15,55					12,17	15,97
	18,29	15,94					12,32	15,56
							12,43	15,25
							12,08	14,96
							11,73	14,69
							10,48	14,37
							14,07	14,35
							19,49	14,69
							16,63	14,81
Mean Elapsed Time:	14,08				12,69		13,67	
Wed, Oct. 31	16,39	15,96			13,49	13,74	15,69	14,83
	21,57	16,12			15,28	13,79	15,37	14,86
	18,79	16,41			13,97	13,87	14,81	14,85
					18,93	13,97	16,18	14,92
							14,25	14,89
							16,97	14,98
							14,34	14,95
							11,69	14,82
Mean Time:	18,58				15,41		14,84	

BEST COPY AVAILABLE

Table 2. (Continued)

Searcher:	AH		JT		JR		DT	
Thurs., Nov. 1	Elapsed Times	Cumulative Mean	Elapsed Times	Cumulative Mean	Elapsed Times	Cumulative Mean	Elapsed Times	Cumulative Mean
			12.50	14.20	14.28	13.98	11.14	14.67
			14.99	14.22	12.21	13.92	13.17	14.61
			13.35	14.19	15.04	13.95	10.92	14.48
			13.40	14.17	13.84	13.95	16.77	14.56
			12.94	14.14	15.23	13.99		
			14.80	14.16	12.44	13.94		
			12.33	14.12	14.27	13.95		
			11.87	14.07	13.12	13.93		
			13.74	14.06				
			12.80	14.03				
Mean Elapsed Time:			13.27		13.80		13.00	

Summary Statistics:

Searcher:	AH	JT	JR	DT	All
Total Number of Searches Completed:	29	46	39	28	142
Mean of All Searches:	16.21	14.03	13.93	14.56	14.55
Sample Standard Deviation:	2.5336	1.5774	2.1652	3.0552	

APPENDIX B
INCREMENTAL SEARCH TOPICS

TOPICS	FACET 1	FACET 2	FACET 3
1. Deaf children and media	deaf (10)*	media (17)	--
2. Joint community and school use of facilities	cooperative programs (3)	facilities (13)	--
3. Resources of Indian reservations	reservations (1)	resources (15)	--
4. Teaching physics by the discovery method	discovery (11)	physics, sciences (8)	teaching techniques (10)
5. Library service to institutions	library service (4)	institutions (18)	--
6. Credit in alternative schools	alternative schools (7)	credit (6)	--
7. Remedial education in high school: evaluation	remedial education (12)	evaluation (6)	high school (SEARCH SAVE)
8. Attitudes toward the environment	attitudes (13)	environment (15)	--
9. Scientific inquiry training in high school	inquiry (9)	science (14)	high school (SEARCH SAVE)
10. Library evaluation	libraries (14)	evaluation (15)	--
11. Library cost accounting	libraries (12)	costs (19)	--
12. Perceptual salience	perceptual development (12)	psycholinguistics (14)	preschool/ primary (14)
13. Preservice programs in science education	preservice education (7)	science education (14)	--
14. Bilingual education in early childhood	bilingual education (11)	early childhood (14)	--

* The number in parentheses is the number of terms used with this facet.

RECENT ILR PUBLICATIONS

Publications of papers and reports of interest to scholars and practitioners in the field of library and information science is an important function of the Institute of Library Research. In addition to this study, the following have been published recently by ILR.

BEST COPY AVAILABLE

ILR-73-001 Todd, Judy, *Summary Report of Student Studies of the Subject Headings Used in the University of California, Berkeley Subject Catalog* (July 1973) 8 pp. (ERIC No. ED-082 775)

ILR-73-002 Bourne, Charles P., and Jo Robinson, *SDI Citation Checking as a Measure of the Performance of Library Document Delivery Systems* (July 1973) 10 pp. (ERIC No. ED-082 774)

ILR-73-003 Weeks, Kenneth, *Determination of Pre-Acquisition Predictors of Book Use: Final Report* (July 1973) 20 pp. (ERIC No. ED-082 776)

ILR-73-004 Weeks, Kenneth, *Proposal for a University of California/California State University and Colleges Inter-Segmental Machine Readable Library Patron Card* (August 1973) 21 pp. (ERIC No. ED-082 777)

ILR-73-005 LeDonne, Marjorie, "Summary of Court Decisions Relating to the Provision of Library Services in Correctional Institutions," *Association of Hospital and Institution Libraries Quarterly* (Winter/Spring 1973) 9 pp.

ILR-73-006 Thelin, John, and Bonnie F. Shaw, (editors), *Institute of Library Research Annual Report: July 1972 to June 1973* (September 1973) 30 pp.

ILR-73-007 Deklevs, Borut, *Uniform Slavic Transliteration Alphabet (USTA)* (October 1973) 82 pp. (ERIC No. ED-086 164)

ILR-73-008 LeDonne, Marjorie, Findings and Recommendations. Volume I, *Survey of Library and Information Problems in Correctional Institutions* (January 1974) 88 pp.

ILR-73-009 LeDonne, Marjorie, Access to Legal Reference Materials in Correctional Institutions. Volume II, *Survey of Library and Information Problems in Correctional Institutions* (January 1974) [illegible] pp.

ILR-73-010 LeDonne, Marjorie, David Christiano, and Jane Scantlebury, Current Practices in Correctional Library Services: State Profiles. Volume III, *Survey of Library and Information Problems in Correctional Institutions* (January 1974) 68 pp.

ILR-73-011 LeDonne, Marjorie, David Christiano, and Joan Stout, Bibliography. Volume IV, *Survey of Library and Information Problems in Correctional Institutions* (January 1974) 28 pp.

ILR-73-012 Gregor, Dorothy, *Feasibility of Cooperative Collecting of Exotic Foreign Language Serial Titles among Health Sciences Libraries in California* (February 1974) 44 pp.

ILR-74-001 Nozik, Barbara, *The Use Status of Books Requested from the University of California, Berkeley, Inter-Library Loan* (March 1974) 11 pp.

ILR-74-002 Bourne, Charles P., *Institute of Library Research Annual Report: July 1973 to June 1974* (1974) 25 pp. + appendices

ILR-74-003 Humphrey, Allan J., *Survey of Selected Installations Actively Searching the ERIC Magnetic Tape Data Base in Batch Mode*. Volume I (June 1973) 86 pp.

ILR-74-004 Cooper, William S., Donald T. Thompson, and Kenneth R. Weeks, *The Duplication of Monograph Holdings in the University of California Library System* (October 1974) 32 pp.